

DIALOG(R) File 351:Derwent WPI  
(c) 2001 Derwent Info Ltd. All rts. reserv.

012613448    \*\*Image available\*\*  
WPI Acc No: 1999-419552/199936  
XRPX Acc No: N99-313212

**Multi media editing method e.g. for creating dynamic production for  
sensory perception by user**

Patent Assignee: CANON KK (CANO ); CANON INFORMATION SYSTEMS RES AUSTRALIA  
(CANO )

Inventor: WINDLE J R

Number of Countries: 002    Number of Patents: 003

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
AU 9888326	A	19990422	AU 9888326	A	19981006	199936    B
JP 11234560	A	19990827	JP 98283153	A	19981005	199945
AU 719329	B	20000504	AU 9888326	A	19981006	200030

Priority Applications (No Type Date): AU 979601 A 19971003

Patent Details:

Patent No	Kind	Lan Pg	Main IPC	Filing Notes
AU 9888326	A	70	H04N-005/222	
JP 11234560	A	112	H04N-005/265	
AU 719329	B		H04N-005/222	Previous Publ. patent AU 9888326

Abstract (Basic): AU 9888326 A

NOVELTY - The method involves extracting from a portion of one element content information relating to a perceivable content of the portion. The production is created by applying a predetermined set of production rules to the portion according to the extracted content information.

DETAILED DESCRIPTION - An INDEPENDENT CLAIM is included for a computerized method for forming a dynamic media production from source media having elements including audio media and visual media, and a media production system.

USE - For multi media editing. For creating multi media productions that uses content information of elements of production to guide in construction of multi media production.

ADVANTAGE - Provides guided creation of production which provides assistance and guidance to user creating such a production.

DESCRIPTION OF DRAWING(S) - The figure shows a schematic representation of an overview of an apparatus in accordance with a preferred embodiment of the invention.

pp; 70 DwgNo 1/28

Title Terms: MULTI; MEDIUM; EDIT; METHOD; DYNAMIC; PRODUCE; SENSE;  
PERCEPTION; USER

Derwent Class: T01; W04

International Patent Class (Main): H04N-005/222; H04N-005/265

International Patent Class (Additional): G06F-017/30; G06F-019/00;

G06T-011/80; G11B-027/02

File Segment: EPI

Manual Codes (EPI/S-X): T01-J; T01-J10C5; T01-J10C7; T01-J30; W04-G; W04-K;  
W04-K10; W04-N05C3; W04-P

**Best Available Copy**

This Page Blank (uspto)

(19) 日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-234560

(43) 公開日 平成11年(1999) 8月27日

(51) Int.Cl.<sup>8</sup>

識別記号

F1

H04N 5/265

H04N 5/265

G06F 17/30

G06F 15/40

370G

G06T 11/80

15/401

330Z

15/62

320K

審査請求 未請求 請求項の数59 OL 外国語出願 (全112頁)

(21) 出願番号 特願平10-283153

(71) 出願人 000001007

キヤノン株式会社

東京都大田区下丸子3丁目30番2号

(22) 出願日 平成10年(1998)10月5日

(74) 代理人 弁理士 太塚 康徳 (外2名)

(31) 優先権主張番号 P09601

(32) 優先日 1997年10月3日

(33) 優先権主張国 オーストラリア (AU)

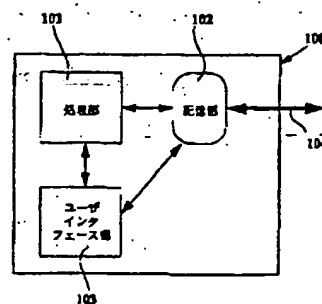
最終頁に続く

(54) 【発明の名称】 マルチメディア編集方法及び装置

(57) 【要約】

【課題】 ユーザーに対して補助や指導を提供するガイド付きの制作物創作のための装置、システム及び/又は方法を提供すること。

【解決手段】 少なくとも1つの音声メディア及び画像メディアを含むソースメディアからメディア作品を制作する方法である。ソースメディアはソースメディアの要素に対応付けされたコンテンツ情報を抽出するために解析される。コンテンツ情報は対応する要素に適用可能な製作規則の限定されたセットを決定するため、少なくとも1つの要素について調べられる。ユーザは要素の中の対応する一つに、要素を例えば持続時間や他の要素との同期性などについて変更するように少なくとも1つの規則を選択的に適用することができる。変更された要素は少なくとも1つの他の要素と結合され、メディア制作物を形成する。



## 【特許請求の範囲】

【請求項1】 コンピュータによって実行される、ユーザによる、知覚的認識のための動的制作物の創作方法であって、前記制作物がユーザが理解できる少なくとも1つの音声要素及び画像要素を含み、前記方法が(a)少なくとも1つの前記要素の少なくとも1つの部分から、その部分の知覚可能なコンテンツに関連するコンテンツ情報を抽出するステップと(b)前記制作物を、抽出されたコンテンツ情報に従って、前記部分に制作規則の所定の(第1)セットの少なくとも1つを適用して制作するステップとを含むことを特徴とする創作方法。

【請求項2】 前記ステップ(a)の前に、(aa)前記少なくとも1つの部分に、制作規則の所定の(第2の)セットの少なくとも1つを適用し、対応するコンテンツ情報を形成するステップと、(ab)前記コンテンツ情報を対応する前記部分に関連づけるステップとをさらに有することを特徴とする請求項1記載の創作方法。

【請求項3】 前記ステップ(b)が前記抽出コンテンツ情報に選択的に適用された前記制作規則の第1セットに基づいた変更部分に相当するような前記部分の編集を含むことを特徴とする請求項1記載の創作方法。

【請求項4】 前記ステップ(b)が、さらに、前記制作規則の第1セットの少なくとも1つを選択的に前記部分に適用し、前記変更部分を形成するために、前記抽出コンテンツ情報及び前記制作規則の第1セットを調べることを特徴とする請求項3記載の創作方法。

【請求項5】 前記要素のそれぞれが前記制作規則の第1及び第2のセットの範囲内の対応する規則を有し、ステップ(aa)が対応する制作規則を前記部分内の対応する前記要素に適用することを特徴とする請求項2記載の創作方法。

【請求項6】 前記要素が少なくとも1つの音声要素、動画像要素及び静止画像要素を含み、前記部分が少なくとも1つの前記要素または前記要素の部分の特徴を有することを特徴とする請求項1記載の創作方法。

【請求項7】 前記要素が音声要素で、前記部分が前記音声要素に適用されるラウドネス、ピッチ、テンポまたはビート及び遅延少なくとも1つである請求項6記載の創作方法。

【請求項8】 前記要素が画像要素で、前記部分が前記画像要素の明るさ、色、動き、フレーム、フレームのセグメントの少なくとも1つである事を特徴とする請求項6記載の創作方法。

【請求項9】 音声メディア及び画像メディア少なくとも1つの要素を有するソースメディアから動的メディアを形成する電算化された方法であって、(a)前記ソースメディアを受信するステップと、(b)前記ソースメディアの少なくとも1つの部分について前記制作規則の所定のセットを用いて前記要素の少なくとも1つを

分析し、そこから前記ソースメディアの対応する部分に関連したコンテンツ情報を抽出するステップと、(c)対応する前記要素に適用可能な制作規則の限定セットを決定するために前記部分の少なくとも1つのコンテンツ情報を調べるステップと、(d)前記対応する要素を変更するように前記制作規則の少なくとも1つを対応する前記要素の1つに選択的に適用するステップと(e)少なくとも前記変更要素を他の前記メディアと結合し、前記メディア制作物を形成するステップとを有することを特徴とする電算化された方法。

【請求項10】 ステップ(b)、(c)及び(e)が実質的に自動的に処理され、ステップ(d)においてユーザが前記一つの規則の選択と、選択された規則が適用される前記一つの要素の対応を示すことを特徴とする請求項9記載の電算化された方法。

【請求項11】 ステップ(c)が選択ステップ(d)を予想して、前記少なくとも1つの要素と、対応する前記規則の限定されたセットをユーザに表示することを特徴とする請求項10記載の電算化された方法。

【請求項12】 前記ソースメディアが音声及び画像要素を含み、前記規則が前記要素それぞれの持続時間を含み、前記方法が、他の音声及び画像要素の希望する持続時間と釣り合うように、前記音声及び画像要素の少なくとも1つの持続時間を修正することを特徴とする請求項9記載の電算化された方法。

【請求項13】 前記ソースメディアが音声要素及び前記音声要素のビートを含む、関連するコンテンツ情報を含むと共に、前記方法が前記ビートに従った前記音声要素の持続時間の変更を含むことを特徴とする請求項9記載の電算化された方法。

【請求項14】 前記変更された持続時間が前記音声要素のビート期間の倍数であることを特徴とする請求項13記載の電算化された方法。

【請求項15】 前記ソースメディアが前記変更された音声要素と同時に再生されることが望まれる画像要素を含み、前記方法が前記変更された音声要素の持続時間と関連した前記画像要素の持続時間のトリミングを含むことを特徴とする請求項12記載の電算化された方法。

【請求項16】 前記画像要素が前記変更された音声要素の持続時間と一致するようにトリミングされることを特徴とする請求項15記載の電算化された方法。

【請求項17】 前記画像要素が連続したフレームから構成され、前記トリミングが前記フレームの選択されたフレームを省略する工程を有することを特徴とする請求項16記載の電算化された方法。

【請求項18】 前記フレームの連続が再生されるべき個々の場面を有し、前記トリミングが前記場面の少なくとも1つの持続時間を、前記場面間のトランジションが前記ビート期間の倍数と一致するように変更することを特徴とする請求項17記載の電算化された方法。

【請求項19】 前記フレームの選択を省略規則のセットに従って調べ、前記連続したフレームからどのフレームを省略するのがもっとも好ましいかを決定するステップを有し、それから決定されたフレームの選択されたものを省略することを特徴とする請求項16記載の電算化された方法。

【請求項20】 前記省略規則のセットが、もし除去された場合に前記メディア制作物の画質に与える影響がもっとも小さくなるようなフレームを定めるように決められていることを特徴とする請求項19記載の電算化された方法。

【請求項21】 前記省略規則が、

(i) 他の前記フレームに関して動きの影響がもっとも少ないフレームの検出と、

(ii) 他のフレームを除去するための、パンまたはズームの中央に近いフレームの検出と、

(iii) 他のフレームに比較してフォーカスが小さくされたフレームを検出し、フォーカスの小さくされたフレームを省略することを含むことを特徴とする請求項20記載の電算化された方法。

【請求項22】 前記画像要素に適合させるために前記音声要素の再生のためのビート期間を変更するステップをさらに有することを特徴とする請求項17記載の電算化された方法。

【請求項23】 前記音声要素の持続時間が画像要素の場面間のトランジションに従って修正され、前記トランジションがバット編集、ワイプ及びフェードを含むことを特徴とする請求項17記載の電算化された方法。

【請求項24】 前記音声要素に対応した音声出力レベルを修正するステップをさらに含むことを特徴とする請求項12記載の電算化された方法。

【請求項25】 前記音声要素の前記コンテンツ情報が前記メディア制作物における代用のための望ましい特性を有する代用音声要素の選択に使用されることを特徴とする請求項9記載の電算化された方法。

【請求項26】 前記特性が平均音量、ビート期間及び優勢な周波数からなるグループから選択されることを特徴とする請求項25記載の電算化された方法。

【請求項27】 前記画像要素の前記コンテンツ情報が前記画像要素の創作時に前記画像要素と共に符号化されたメタデータを含むことを特徴とする請求項9記載の電算化された方法。

【請求項28】 前記コンテンツ情報が前記画像要素の中に存在する画像に関するズームデータを含むことを特徴とする請求項27記載の電算化された方法。

【請求項29】 前記コンテンツ情報を含む前記画像要素の特性に基づき前記画像要素を代表する画像を印刷するステップをさらに含み、前記印刷された画像が前記ユーザに対して前記画像要素の重要部分の目印を形成することを特徴とする請求項9記載の電算化された方法。

【請求項30】 前記画像要素が前記画像要素の色及び/又は輝度に関連したコンテンツ情報を含むことを特徴とする請求項9記載の電算化された方法。

【請求項31】 色コンテンツ情報に基づき、画像要素からユーザ生成画像要素を生成するステップを更に有することを特徴とする請求項30記載の電算化された方法。

【請求項32】 前記ユーザ生成画像要素が少なくとも1つのテキスト要素とグラフィックオブジェクトを有することを特徴とする請求項31記載の電算化された方法。

【請求項33】 前記コンテンツ情報が前記ソースメディアの創作とともに生成されることを特徴とする請求項9記載の電算化された方法。

【請求項34】 前記ステップ(b)がステップ(a)と実質的に同時に実行されることを特徴とする請求項9記載の電算化された方法。

【請求項35】 前記コンテンツ情報が前記画像要素の隣接する部分間におけるワイプ又はディゾルブの決定に用いられることを特徴とする請求項30記載の電算化された方法。

【請求項36】 請求項1乃至35のいずれかに記載の方法を実行するように構成された装置。

【請求項37】 請求項1乃至35のいずれかに記載の方法を実行するように配列されたコンピュータ実施可能なステップの列が組み込まれた、コンピュータが読み込み可能な媒体を含むコンピュータプログラム製品。

【請求項38】 少なくとも1つの音声メディア及び画像メディアを含むソースメディアを受信するためのメディア入力配列と、

前記ソースメディアから対応する要素に関連づけられたコンテンツ情報を抽出するために前記ソースメディアを解析する解析手段と、

対応する前記要素に適用可能な制作規則の限定されたセットを決定するため、少なくとも1つの前記要素のコンテンツ情報を調べる調査手段と、

前記システムのユーザに、少なくとも前記制作規則のセット及び対応する前記要素の表現を表示し、対応する前記要素に適用すべく前記規則から1つを選択するユーザ入力命令を受信する命令手段と、

前記対応する1つの要素を変更するように、前記選択された規則を対応する前記要素の1つに適用する修正手段と、

少なくとも1つの前記変更された要素と他の前記メディアとを結合し、前記メディア制作物を形成する結合手段とから構成されるメディア制作システム。

【請求項39】 前記ソースメディアが音声及び画像要素を含み、前記規則が前記要素のそれぞれの持続時間を含むとともに、前記音声及び画像要素の少なくとも1つの持続時間が前記音声及び画像要素の他の一方の持続時

間に合致するように修正可能であることを特徴とする請求項38記載のメディア制作システム。

【請求項40】 前記ソースメディアが音声要素を含み、前記関連づけられたコンテンツ情報が前記音声要素のビートを含むとともに、前記修正手段が前記ビートに従って前記音声用その持続時間を変更するように動作することを特徴とする請求項38記載のメディア制作システム。

【請求項41】 前記変更された持続時間が前記音声要素のビート期間の倍数であることを特徴とする請求項40記載のメディア制作システム。

【請求項42】 前記ソースメディアが前記変更された音声要素と同時に再生されることが望まれる画像要素を含み、前記修正手段が前記変更された音声要素の持続時間と関連した前記画像要素の持続時間をトリミングすることを特徴とする請求項41記載のメディア制作システム。

【請求項43】 前記画像要素が前記変更された音声要素の持続時間と一致するようにトリミングされることを特徴とする請求項42記載のメディア制作システム。

【請求項44】 前記画像要素が連続したフレームから構成され、前記トリミングが前記フレームの選択されたフレームの省略を有することを特徴とする請求項43記載のメディア制作システム。

【請求項45】 前記フレームの連続が再生されるべき個々の場面を有し、前記トリミングが前記場面の少なくとも1つの持続時間を、前記場面間のトランジションが前記ビート期間の倍数と一致するように変更することを特徴とする請求項44記載のメディア制作システム。

【請求項46】 前記調査手段が、前記フレームの連続を省略規則のセットに従って調べ、前記連続するフレームからどのフレームを省略するのがもっとも好ましいかを決定し、前記修正手段が前記ユーザの選択に応じて決定されたフレームのフレームを省略するように動作可能であることを特徴とする請求項44記載のメディア制作システム。

【請求項47】 前記省略規則のセットが、もし除去された場合に前記メディア制作物の画質に与える影響がもっとも小さくなるようなフレームを定めるように決められていることを特徴とする請求項46記載のメディア制作システム。

【請求項48】 前記省略規則が、

(i) 他の前記フレームに関して動きの影響がもっとも少ないフレームの検出

(ii) 他のフレームを除去するための、パンまたはズームの中央に近いフレームの検出

(iii) 他のフレームに比較してフォーカスが小さくされたフレームを検出し、フォーカスの小さくされたフレームを省略することを含むことを特徴とする請求項46記載のメディア制作システム。

【請求項49】 前記修正手段が、前記画像要素に適合させるために前記音声要素の再生のためのビート期間を変更するように動作可能であることを特徴とする請求項41記載のメディア制作システム。

【請求項50】 前記修正手段が、前記音声要素の持続時間を画像要素の場面間のトランジションに従って修正するように動作可能であり、前記トランジションがバット編集、ワイプ及びフェードを含むことを特徴とする請求項40記載のメディア制作システム。

【請求項51】 前記修正手段が、前記音声要素に対応した音声出力レベルを修正するように動作可能であることを特徴とする請求項40記載のメディア制作システム。

【請求項52】 前記音声要素の前記コンテンツ情報が前記メディア制作物における代用のための望ましい特性を有する代用音声要素の選択に使用されることを特徴とする請求項38記載のメディア制作システム。

【請求項53】 前記特性が平均音量、ビート期間及び優勢な周波数からなるグループから選択されることを特徴とする請求項52記載のメディア制作システム。

【請求項54】 前記画像要素の前記コンテンツ情報が前記画像要素の制作時に前記画像要素と共に符号化されたメタデータを含むことを特徴とする請求項38記載のメディア制作システム。

【請求項55】 前記コンテンツ情報が前記画像要素の中に存在する画像に関するズームデータを含むことを特徴とする請求項54記載のメディア制作システム。

【請求項56】 前記コンテンツ情報を含む前記画像要素の特性に基づき前記画像要素を代表する画像を印刷する手段をさらに含み、前記印刷された画像が前記ユーザに対して前記画像要素の重要部分の目印を形成することを特徴とする請求項38記載のメディア制作システム。

【請求項57】 前記画像要素が前記画像要素の色に関連したコンテンツ情報を含むことを特徴とする請求項38記載のメディア制作システム。

【請求項58】 色コンテンツ情報に基づき、画像要素からユーザ生成画像要素を生成する手段を更に有することを特徴とする請求項57記載のメディア制作システム。

【請求項59】 少なくとも1つの音声メディア及び画像メディアを含むソースメディアから動的メディア制作物を形成するために配置されたコンピュータ実施可能なステップの列が組み込まれたコンピュータ読み込み可能な媒体であって、前記コンピュータ実施可能なステップの列が、

前記ソースメディアを受信するための第1モジュールと、

前記ソースメディアの1部の対応する要素に関連づけられたコンテンツ情報を抽出するため、予め定めたコンテンツ規則のセットを用いて前記ソースメディアの少なく

とも前記1部に対して少なくとも1つの要素を解析する第2モジュールと、  
対応する前記要素に適用可能な制作規則の限定されたセットを決定するために少なくとも1つの前記1部のコンテンツ情報を調査する第3モジュールと、  
前記要素の対応する1つを変更するように、前記対応する1つの要素に少なくとも1つの前記制作規則を選択的に適用する第4モジュールと、  
少なくとも1つの前記変更された要素とその他の前記メディアを結合し、前記メディア制作物を形成する第5のモジュールとを含むことを特徴とする媒体。

#### 【発明の詳細な説明】

#### 【0001】

【発明が属する技術分野】本発明は全体としてマルチメディア編集に関し、特に構成要素のコンテンツ情報を製品構築におけるガイドに用いるマルチメディア製品の創作方法に関する。

#### 【0002】

【従来の技術】静止画、文字、ビデオ、音及び他のマルチメディア要素をまとめて一つの製品にするのは、多様な領域における膨大な知識を必要とする複雑な作業である。多種の要素をまとめて一つの最終制作物にするということは、ビデオ編集において既に十分確立した領域である。ビデオ制作物においてしばしば用いられる要素の種類は画像、音声及び文字である。画像要素はビデオ、静止画及びアニメーション等を含む。音声要素は背景音、音声効果、コメント及びビデオクリップの録音音声等を含む。文字要素はタイトル、クレジット及びサブタイトル等を含む。

【0003】これらの要素は変更したり、効果を付けたりすることのできるいろいろな特性を有している。そして、要素は最終制作物を形成するために多くの異なる方法で結合することができ、それぞれの結合において行われなければならない選択に関する多くの知識が存在する。

【0004】長年、このような制作はテレビ局や大規模な制作会社でのみ可能だった。家庭用ビデオカメラ、パーソナルコンピュータ、安価なビデオキャプチャカード及びビデオ編集ソフトウェアの出現によって、現在では多くの人々が自分のビデオ制作物を創作することが可能となっている。

【0005】しかしながら、よい選択を行うのに必要な知識は、制作物の制作者から教わらなければならない。2つか3つの単純な制作物を作りたいと考えている多くの人にとって、学習曲線は非常に急なものであり、結果として知識を用いた判断が行われなかったことによる低品質の制作物が創作されている。デジタル化された静止画と音楽、静止画と文字、ビデオ素材単独または音楽素材単独から構成される制作物を創作しようとした場合にも、同様の問題が起る。

【0006】よって、ユーザーに対して補助や指導を提

供するガイド付きの制作物創作のための装置、システム及び/又は方法に対する必要性が存在する。

#### 【0007】

【課題を解決するための手段】本発明の第1の要旨は、コンピュータによって実行される、ユーザによる、知覚的認識のための動的制作物の創作方法であって、制作物がユーザが理解できる少なくとも1つの音声要素及び画像要素を含み、方法が(a)少なくとも1つの要素の少なくとも1つの部分から、その部分の知覚可能なコンテンツに関連するコンテンツ情報を抽出するステップと、  
(b)制作物を、抽出されたコンテンツ情報に従って、部分に制作規則の所定の(第1)セットの少なくとも1つを適用して創作するステップとを含むことを特徴とする創作方法に存する。

【0008】また、ステップ(a)の前に、(aa)少なくとも1つの部分に、制作規則の所定の(第2の)セットの少なくとも1つを適用し、対応するコンテンツ情報を形成するステップと、(ab)コンテンツ情報を対応する部分に関連づけするステップとをさらに有することが好ましい。

【0009】また、本発明の別の要旨は、音声メディア及び画像メディア少なくとも1つの要素を有するソースメディアから動的メディアを形成する電算化された方法であって、(a)ソースメディアを受信するステップと、(b)ソースメディアの少なくとも1つの部分について制作規則の所定のセットを用いて要素の少なくとも1つを分析し、そこからソースメディアの対応する部分に関連したコンテンツ情報を抽出するステップと、

(c)対応する要素に適用可能な制作規則の限定されたセットを決定するために部分の少なくとも1つのコンテンツ情報を調べるステップと、(d)対応する要素を変更するように制作規則の少なくとも1つを対応する要素の1つに選択的に適用するステップと(e)少なくとも1つの変更要素を他のメディアと結合し、メディア制作物を形成するステップとを有することを特徴とする電算化された方法に存する。

【0010】装置、システム及びコンピュータプログラム製品を含む本発明の他の見地は以下の説明及び本明細書に記載された請求の範囲を理解することによって明らかになるであろう。

#### 【0011】

【発明の実施の形態】図1は、マルチメディア編集装置100の好ましい実施形態の概要を示すブロック図である。マルチメディア編集装置100は処理部101、記憶部102、ユーザーインタフェース部103及び通信ポートとそれに対応するバス104とを有している。記憶部102には処理部101がユーザーインタフェース部103のコントロール及びユーザーインタフェース部103に対する応答のために用いるアプリケーションプログラムコードが記憶されている。記憶部102はま

た、制作物の制作に使われる種々の規則や、制作物に使用される1つ以上の要素の保存にも使用される。ここにおいて、「制作物」という単語は一般的な意味で用いられている。それは制作物がビデオ制作物、音楽制作物、スライドショーの様に表現される連続した静止面写真やそれらの組み合わせとなりうるからである。

【0012】上述の通り、要素はきわめて一般的に、制作物の構成要素として表現することのできる、画像要素、音声要素及び／又は文字要素を含む。画像要素はビデオ(video footage)及びその一部であるビデオクリップ、1ビデオフレーム、映画全体、コンピュータクリップアート、写真及びビデオアニメーションで用いられるような作られた画像を含んでよい。音声要素は音声記録、音声効果、音楽及び楽譜、一つ及び多重の調子又は音色を含んでよい。文字要素は制作物のタイトル及びクレジット、ロゴやスローガンのような広告及び／又は他のしるしを含んでよい。

【0013】好ましい実施形態の概要において、通信ポート及び対応するバス104は記憶部102へのデータ転送及び記憶部102からのデータ転送に用いられる。このデータは典型的には上述した要素のデジタル表現である。通信ポート及び対応するバス104は例えばアプリケーションソフトウェアコードのダウンロード又はアップロードなど他の目的にも使用される。バス104及び対応するインターフェースは、例えばIEEE1394高速シリアルバス標準で規定されたものである。

【0014】処理部101はコンテンツ情報を抽出するために、アプリケーションソフトウェアに補助されて要素を解析する。抽出されたコンテンツ情報は記憶部102に保存され、そのコンテンツ情報を抽出した要素と対応付けされる。コンテンツ情報は、多種の異なる種類の要素(例えば画像、音声、文字)に固有のもので、処理部101がソフトウェアと連携して行う上述の解析も適宜変化する。多種の要素に対し、コンテンツ情報の抽出(いくつかの場合にはその生成)のために行われる解析については、以下の対応する部分において説明されるであろう。

【0015】コンテンツ情報を抽出する解析は、多種の要素が記憶部102にダウンロードされてからある時間後になされる。一つのオフラインプロセスとして行うことができる。しかし、全部ではない少なくともいくつかの解析を、バス104から記憶部102にダウンロードされた多種のデータに対し、その場で(オン・ザ・フライで)行うことが可能であることは想像される。好ましい構成において、多種のデータに対する解析は少なくともその一部分を、要素の獲得に用いられる録音又はキャプチャ装置が行う。例えば、デジタルビデオカメラ(DVC)は、画像及び音声データだけでなく、日付、時間、ビデオフレームの数等のメタデータの保存/記録を行うように改造することができる。好ましい実施形態に

おい用いることができる。選択された要素をキャプチャリングするDVCの例は、以下「拡張収集装置」(Enhanced Gathering Apparatus)として説明される。

【0016】図2は、制作物を生成する複数の画像要素と音声要素からなる入力スクリプト1300を示すブロック図である。図2の入力スクリプト1300はビデオプレゼンテーション制作物の例を示すが、音声制作物、スライドショータイプの静止画プレゼンテーション、印刷物としての制作物(本、カタログ等)等、他の形式に対する入力スクリプトもまた可能である。スクリプト1300の画像要素はビデオセグメント1302、1304及び1305を有している。これらセグメントはデジタル形式で保存されていることが好ましい。また、デジタル静止画1303は、ディスプレイ上で静止画としての見かけを提供するように、ある時間再生された一つの画像や、予め定めた期間ポーズされた1フレームの画像を含むビデオセグメントでもよい。

【0017】音声要素は音声バックイングトラック1301及び1306から構成される。一般的に、音声バックイングトラックは音楽トラック及び／又は音声効果を有する。また、いくつかの場合、音声バックイングトラックはビデオセグメントに対応した音声トラックに付加される音声トラックである。この説明を通して、逆の明示をしない限り、ビデオセグメントという参照は、対応する有効な音声トラックを含んだビデオセグメントを示す。例えば、ビデオには通常、人の声や背景ノイズといった、それに対応した録音音声がある。録音音声の情報は好ましい実施形態の説明の目的において、ビデオセグメントの一部であると仮定するが、実施形態の実施は、画像及び音声源情報のいずれかのみを用いることが好ましい。

【0018】入力スクリプトは、状況に応じて、または必要に応じて順次又は並行して実行(再生)される。入力スクリプト1300は好ましくは順次実行される。しかし、大半のマルチメディア装置において、概して音声要素は異なるチャンネルを介して並行して画像要素に供給される。入力スクリプト1300の配列において、第1音声バックイングトラック1301が開始し、画像要素1302、1303及び1304の表示装置への出力と同時に、音声装置に出力される。画像要素1304の終わりに、音声バックイングトラック1301から次のシーケンシャル音声バックイングトラック1306へのトランジションが起こり、異なるシーケンシャル画像要素1305が音声バックイングトラック1306と同時に表示される。画像用途と音声バックイングトラックとの同期が望まれ、要素が画像、音声、文字のいずれであっても、各要素とそれに関連するタイミング情報が符号化されていることが好ましい。例えば、画像及び／又は音源の録音中のように、コンテンツ情報の抽出はライブで行われる。これはユーザーの入力によって自動又は半自動で実現される。



【0019】いくつかの例において、コンテンツ情報は音終了後の手動制御または自動化された後処理によって生成される。各要素は更に、

- ・処理アプリケーションに元データ及び要素に対応したコンテンツデータを示す情報
  - ・要素の期間、元データに関するオフセットスタート時間等の捕捉情報
- を有することが好ましい。

【0020】図3は、好ましい実施形態におけるビート同期プロセス1420のフローチャートである。ビート同期プロセス1420は、予め定められたシーケンス又は順序で、複数の要素、すなわち要素の入力アレイの形で入力を受信する。ビート同期プロセス1420の出力は、図2に示す入力スクリプト1300の形式と同じ入力スクリプトである。しかしながら、要素の入力アレイを構成する各要素は、ビート同期プロセス1420において変化したりしなかったりする。

【0021】まず、ステップ1400において、ビート同期プロセス1420は入力として、制作物の要素（画像、音声、その他）のアレイを、好ましくは予め定めた順序で受信する。次に、入力アレイの要素の数と実質的に似た数の入力スロットを有する空の出力アレイがステップ1401において作られる。この作成ステップ1401において、ポインタもまた入力アレイの第1要素の開始点に初期化される。コントロールは次のステップである、入力アレイ中に処理すべき要素があるかどうかを決定する決定ステップ1414に移行する。ステップ1414の決定結果が「偽」であれば、現在の出力アレイが入力スクリプト1300で示される形式で出力される。一方、少なくとも1つの要素が存在した場合には、決定ステップ1414は「真」を返し、入力アレイ（ソースアレイ）の現在の要素を出力アレイのスロットにコピーするコピーステップ1402に移行する。

【0022】フローチャート1420において、コピーステップ1402の次にある音声要素決定ステップ1403では、現在の要素が音声バックトラックであれば真を、そうでなければ偽を返す。音声要素決定ステップ1403において真の結果が得られると、プロセスは読み込み入力ステップ1406に移行する。ステップ1406では、現在の要素の補足情報部分からビート期間を読み取り、それに応じてカレントビート期間をセットする。このステップではビート期間の計算を含んでもよい。ポインタはステップ1413において入力アレイの次の要素に進む。音声要素決定ステップ1403において偽の結果が得られた場合には、追加の決定ステップ1404においてカレントビート期間が既にセットされているかどうかを調べる。決定ステップ1404は、カレントビート期間が既にセットされていれば真を、そうでなければ偽を返す。ステップ1404の結果が偽の場合、ステップ1413においてポインタが入力アレイの

次の要素に進む。

【0023】決定ステップ1404の結果が真の場合、プロセス1420は更に別の決定ステップ1405へ移行し、ここで現在の要素がカレントビート期間の影響を受けているかどうかを決定する。この決定において現在の要素がカレントビート期間の影響を受けていなければ、ステップ1413でポインタが次の要素に進む。現在の要素がビート期間の影響を受けているかどうかを決定する決定ステップ1405において真の結果が得られた場合には、現在の要素が静止画シーケンスであるかどうかを決定する別の決定ステップ1410に移行する。この決定ステップにおいて、現在の要素が静止画である場合には、トリミングステップ1408が現在の要素に実施され、現在の要素は出力アレイの現在の要素、すなわちステップ1402においてコピーされた要素となる。トリミングステップ1408に続くステップ1413において、ポインタは次の要素に進む。

【0024】決定ステップ1410において偽の結果（すなわち、現在の要素が静止画でない）が返された場合には、決定ステップ1411において、バックトラック又は静止画でない現在の要素が、ズームショットシーケンス、パンショットシーケンス又は他の予め定めたビデオ特殊効果のような、予め定めたビデオコンディションを含むかどうか決定される。現在の要素にいかなるビデオコンディションも検出（決定）されなければ、現在の要素にトリミングステップ1408が実行される。それぞれの場合で要素（画像要素）に対してトリミングステップ1408が実行され、要素はカレントビート期間の実質的な倍数にトリミングされる。

【0025】前述の通り、トリミングステップ1408に続くステップ1413でポインタが次の要素に進む（出力アレイにコピーされるべき次の入力要素がもしあれば）、ステップ1411においてビデオコンディションが確認された場合には、ステップ1411から真が返され、ステップ1412において現在の要素は各ビデオコンディションに対して予め定義された規則に従ってトリミングされる。適用される予め定義された規則の例は後に説明する。トリミングステップ1412に続くプロセスフローは前述したポインタ進行ステップ1413である。ポインタが次の要素に進むと、決定ステップ1407が入力アレイに更なる要素があるかどうかをチェックする。もしあれば、プロセスはコピーステップ1402にループバックし、更なる入力要素をチェックする決定ステップ1407において偽が返され、ステップ1409において出力アレイが出力されてビート同期プロセス1420が再呼び出しされるまで終了する状態になるまで、上述したコピーステップ1402に引き続くステップを実行する。上の説明においては、ビデオをトリミングする場合を説明したが、類似の処理をビデオをビート周波数に適合するように行う伸長又は圧縮に適用して

もよい。

【0026】図4は、好ましい実施形態によるトランジション生成プロセス1500のフローチャートを示す。トランジション生成プロセス1500はビデオ要素間で用いられるトランジションのタイプ及びトランジション期間を決定する。トランジションは最終制作物が出力表示装置に表示されるまでは表現されない。

【0027】最初に、トランジション生成プロセス1500は入力として入力用アレイを受信する。入力アレイの各要素は、対応する1つ又は複数のチャンネルの情報を有する。まずプロセス1500の最初のステップ1501において、入力アレイのエレメント数と実質的に同じ数のスロットを有する空の出力アレイを作成する。このステップ1501において初期化手順も実行される。例えば、ソースポイントが入力アレイの最初の要素を指すようにセットされ、行き先ポイントが出力アレイの最初の有効なスロットを指すようにセットされ、ビート期間

に対する新しい値が決定されるまで、ビート期間変数がデフォルト値にセットされる。さらに、各ビデオチャンネルに対し、変数(すなわち"PrevEle")が先のビデオ要素を示すように支持される。最初は先立つビデオ要素がないので、変数PrevEleは先のビデオ要素がないことを示す値を割り当てられる。初期化に続いて、プロセス1500は決定ステップ1507に移行する。このステップでは入力アレイ中に処理すべき要素があるかないかを決定する。処理すべき要素がなければ、決定ステップ1507は偽を返し、出力アレイの現在の値が図2に示す入力スクリプトの形式で出力される。

【0028】入力アレイが空の場合は、空の入力スクリプトが得られる。すなわち、いかなる入力スクリプトも得られない。しかし、入力アレイに処理すべき要素がある場合には、決定ステップ1507は真を返し、コピーステップ1502に移行する。コピーステップ1502において、現在の入力(アレイ)要素は出力アレイの有効スロットにコピーされる。コピーステップ1502に引き続き、現在の要素(コピーされた要素)が音声バックキントラックであるか否かを決定する決定ステップ1503が実行される。現在の要素が音声バックキントラックならば、現在のビート期間を抽出された値にステップ1506でセットし、ステップ1518において、入力アレイの次の要素にインクリメントされる。

【0029】一方、決定ステップ1503において現在の要素が音声バックキントラックでないと決定された場合は、次のステップ1504において、現在の要素にトランジションを必要とする先のビデオ要素が現在の要素のビデオチャンネルにあるかないかを決定する。もし現在のビデオチャンネルに先立つビデオ要素がなければ、変数PrevEleを現在の要素を指すようにセットするPrevEleセットステップ1505が実行され、インクリメントステップ1518へ移行する。一方、先立つビデオ要素(現

在の(ビデオ)要素に対する)が存在する場合は、決定ステップ1508に移行し、ここでマルチメディア編集装置100のユーザが固定トランジションスタイル(例えばバットエディットトランジション)を選択しているか否かをチェックする。固定トランジションが選択されている場合は、マーキングステップ1510において、現在のビデオ要素へトランジションするために先のビデオ要素にマークする。

【0030】固定トランジションが選択されていない場合、決定ステップ1508から色データ抽出ステップ1513に移行する。輝度データは先のビデオ要素の終端及び現在のビデオ要素の始端から抽出されることが好ましい。次のステップ1514において、抽出された色データに基づき、ステップ1515で選択されるディゾルブトランジションか、ステップ1516で選択されるワイプトランジションのいずれかを決定する。ワイプトランジションの場合、さらなるステップ1517において、ワイプトランジションの詳細を決定するための、先のビデオ要素と現在のビデオ要素の動きベクトル解析を行う。ワイプトランジションの詳細には、開始点、ワイプ速度等が含まれる。

【0031】ワイプトランジションの選択による追加ステップ1517又はディゾルブトランジション選択ステップ1515からはいずれも前述のマーキングステップ1510に戻る。マーキングステップ1510において、先のビデオ要素は現在のビデオ要素へのトランジションのために、選択されたトランジション(すなわち、ディゾルブ、ワイプ、固定、等)に応じてマークされる。マーキングステップ1510に続き、決定ステップ1511において選択されたトランジションが予め定めた持続時間よりも長い持続時間を有するか否かがチェックされる。決定ステップ1511においてはまた、所定持続時間に加え、或いはその替わりに固定トランジションのチェックを行っても良い。

【0032】決定ステップ1511において、"真"値は選択されたトランジションの持続時間が予め定めた持続時間よりも長いことを示し、その場合にはステップ1512において、選択されたトランジションの持続時間を現在のビート期間の倍数となるようにマークする。一方、決定ステップ1511が"偽"を返した場合は選択されたトランジションの持続時間が予め定めた持続時間よりも短いことを示す。この場合には、PrevEleセットステップ1505に戻る。さらに、決定ステップ1512が選択されたトランジションの持続時間にマークすると、同様にPrevEleセットステップ1505に復帰する。

【0033】前述したように、PrevEleセットステップ1505は、変数PrevEleを現在のビデオ要素を指すようにセットし、インクリメントを行う次の入力要素ステップ1518に戻る。ステップ1518はソースポイン

データをインクリメントし、これまで説明したように次の入力要素への処理を継続する。

【0034】トランジション生成プロセス1500は処理する入力要素がなくなった時点で終了する。その場合、出力アレイの現在の要素は図2に示した入力スクリプト1300の形式で出力される。

【0035】図5は、本発明の好適な実施形態による編集プロセス1100の概要を示す。最初に、制作物セットアップステップ1101がユーザに対し、制作物に用いたい画像、音声及び／又は文字要素の選択を許可する。次に、スクリプトプロセス1102がユーザの選択した要素に適用される。これらのスクリプトプロセス1102はそれぞれ図3及び図4を用いて説明したビート同期プロセス及びトランジション生成プロセスを含む。本発明の範囲及びその精神を離れることなく、他のスクリプトプロセス、例えば、静止画又はビデオクリップのフレームのピクチャ要素を操作して編集する編集プロセスを適用することができる。スクリプトプロセスの出力は、実質的に図2を用いて説明した入力スクリプトであることが好ましい。

【0036】スクリプトプロセス1102で得られた入力スクリプトは、次のステップ1103において、スクリプトインタプリタによってそれぞれ翻訳される。決定ステップ1104では、各入力スクリプトに対し、その出力が出力装置で処理されるべきものか否かを決定する。ステップ1104での結果がNOであれば、図5に示されるように、ユーザの更なる選択のために制作物セットアップステップ1101に戻る。しかし、ステップ1104での結果が、出力が出力装置で処理されるべきであることを示すYESであれば、制作物の要素を構成する各入力スクリプトはステップ1105において対応する出力装置において処理される。例えば、音声要素とビデオ要素とからなる入力スクリプトにおいて、音声要素は音声出力装置で演奏し、ビデオ要素に対応したサウンドトラックもまた音声出力装置で演奏しながらビデオ要素のビデオ部分は表示装置で表示する。図5においては単独のインタプリタを参照しているが、一般にはソースデータの各タイプ及び対応する出力装置（例えば音声、画像）に対して1つずつインタプリタが設けられる。

【0037】最後の決定ステップ1106は、制作が終了したのか、引き続きユーザへの選択を許可するために制作セットアップステップ1101に戻るのかをチェックする。セットアップステップへ戻ることにより、ユーザは制作物のプレビューができる。制作物セットアップステップ1101において、ユーザはまたどのスクリプトプロセスを適用するか、或いはしないかについての選択をすることもできる。例えば、ビート同期を行いたくないユーザはそのオプションをオフにする。更に、ユーザは希望する出力装置を選択することもできる。

【0038】図6は、図5におけるスクリプト適用ステップ1102の詳細を示す。最初のステップ1201において、オリジナルの変更を避けるため、ユーザが選択した要素のコピーが行われる。そして、適用されるスクリプトプロセスのリストが作成される。次に、決定ステップ1202がリストにある全てのスクリプトプロセスの適用が終わったか否かを判断する。すなわち、決定ステップ1202はスクリプトプロセスのリストがまだ適用していないスクリプトを有するか否かを判断する。全てのスクリプトプロセスの適用が終了し、結果がNOであれば、スクリプトプロセスで作成された入力スクリプトは図5を用いて説明したようにスクリプトインタプリタに送られる。しかし、ステップ1202においてまだ全ての（リスト上の）スクリプトプロセスの適用が終わっていない場合、コピーされた要素は対応するスクリプトプロセス1203へ引き渡される。例えば、もし現在のコピーされた要素にビート同期プロセスを適用すべき時は、その要素はビート同期プロセスへ向けて送られる。次のステップ1204は適用したスクリプトプロセスからマークを外し、適用すべき他のスクリプトプロセスの有無を調べるため決定ステップ1202へループバックする。適用すべきスクリプトプロセスが無くなると、最後に生成された入力スクリプトが処理1103のためにスクリプトインタプリタに送られる（図5参照）。

【0039】ここで、要素の集まりからの制作物制作をガイドするための、予め定めた規則のセットを定義して良いが、ここで述べる規則のセットは単に例示のために提示されるものであって、それに限定されるべきものではない。その規則のセットは以下に説明される好ましい実施形態に関連して用いることができるが、規則の種類は達成しようとする最終制作物効果に依存する。よって、当業者が認識する、規則の異なる組み合わせ及び／又は異なる種類の規則の適用は、本発明の範囲又は本質から離れることなく適用することができる。例えば、ビデオクリップ全体の動き平均が最も少ないビデオスクリーンの一部に、ビデオクリップのオーバーレイとして、ビデオスクリーンに表示すべきタイトル又はテキスト情報を指定する規則を作成することができる。規則の種類もまた、多種多様の要素から抽出されるコンテンツ情報の種類に依存する。以下のセクションにおいて、制作物の創作そのもの、或いは創作を補助するために使用される、いろいろな要素から抽出された異なるコンテンツ情報及び対応する規則について説明する。

【0040】＜ビート同期＞図7は、ビート210がその上に表された音声トラック200と、ビデオクリップ（又はビデオ）214との自動ビート同期例を示す。一般的に、ビートが検出されてビデオクリップ214との同期に使用される音声トラック200は音楽である。ビートは音声トラックで演じられる音楽の音楽的なビート

(musical beat)であることが好ましいが、音声トラック200から抽出された実質的に一定な周波数も同期目的のビートとして用いることができる。

【0041】ビデオクリップの同期を実現するビートの抽出技術の例を説明する。制作物の音声要素を表現する入力音声信号及び特に音声信号のビートが要求される部分は、音声信号の周波数を所定の範囲に制限するためにバンドパスフィルタを通して分析される。周波数を制限された信号はそのリードエッジを検出するリードエッジ検出回路を通して、ビートを求める音声信号の持続時間において検出された各リードエッジの時間間隔を記録し、平均時間間隔を計算する。この平均時間間隔は音楽を所望のビデオクリップに同期させるビートとして用いることができる。

【0042】図7のビデオクリップ214は、音声トラックのビートと同期を取るべくトリミングされる前の、トリミングされていないビデオを示している。本例において、未トリミングのビデオ214はビデオ214の所定のシーンを表現する3つのセクション201、202及び203から構成される。本例において、シーン間のトランジションはシーンをバットエディティング (butt editing) して構成している。すなわち、通常それぞれが複数のフレームからなる3つのセクションは、端と端で結合されている。

【0043】未トリミングのビデオのセクション201、202及び203間のトランジションは、図7に示された例において、垂直な線204及び205で示されている。未トリミングビデオ204のトランジション204は音声トラック200のビート210と一致していない。未トリミングビデオ214の最初のセクション201は、次のセクション212との間に音声トラック210のビートと一致するトランジション203を有するようなセクション211にトリミングされる。同様に、残りのセクション202及び203はそれぞれセクション212および213にトリミングされ、セクション211、212、及び213の間のトランジションが音声トラックのビート210と同期する。トリミングされたセクション211、212及び213は協調して、シーン中のイベント又は変化の間のトランジションが音声トラック200のビートと同期したビデオクリップ (又はビデオ) を形成する。

【0044】図3のステップ1413で説明したように、ビデオをトリミングし、トリミングされた各セクションからビデオフレームを除去する際には、セクションのどの部分からフレームを除去しても良い。しかし、セクションの内容を所定の効果について評価し、最終的に制作物を見た際の断絶が最小限になるようにフレームを除去することが好ましい。例えば、ズーム又はパンを含むビデオセクションであれば、ズーム又はパンの中間でないフレームを除去することが望ましい。また、ビデオの

動きの少ない部分、すなわち、ビデオの中の本質的に静止面である部分を評価し、動きの少ない部分から実質的にフレームを除去することもできる。さらに、ビデオの所定セクションにおけるビートの合っていない部分を取り除くこともできる。

【0045】さらに、ビデオ214を音声トラック200に同期させる際、音声トラック200とビデオ214の同期精度を向上するために、実質的に音声信号がわずかな程度でビートの時間間隔を増減する処理を音声トラックに行うことを許してもよい。前述したように、編集することなくビデオのスピードを増減するように処理を行っても良い。

【0046】トランジション速度及びタイミング規則一般にビデオフレームレート (フレーム/秒) はほぼ一定であるから、ビデオにおける所定のフレーム数はほぼ特定の時間間隔を規定する。よって、ビデオにおける第1セクションと第2セクション間の時間間隔、すなわちトランジション速度はトランジションに使われるビデオのフレーム数に依存する。トランジションは、図7に示されるような、2つのセクションをバットエディットすることにより得られる。ビデオのあるシーンから異なるシーンへの単純なトランジションでよい。しかし、トランジションは1つのシーンから次のシーンへディゾルブトランジションするといった特殊効果を供給するよう、目的を持って計画することが好ましい。ワイプやフェードを含む他のトランジション形式によって所定の特殊効果を用いてシーンを変えることができる。

【0047】好ましい実施形態において、このようなトランジションのタイミング及びスピードは、予め定めた音声トラックと同期されており、ビデオフレームを複製、付加及び除去できることを除けば、ビート同期について説明した方法と実質的に似た方法でスムーズに同期したトランジション効果が得られる。タイミング及びトランジションスピード規則は最初の所定のビートマーカーにおいて選択したトランジションが開始するように適用される。そして、トランジションを有効にするための所定フレーム数を許可することによって、トランジションの最終フレームが第2の所定のビートマーカーと一致していることを確認する。音声トランジションは交差区間でそれぞれのビートが一致するようにアレンジすることによって達成される。

【0048】音声レベルの調整図8に、音声バックキックトラック301の音声レベル情報と、この音声バックキックトラックと連携しようとするビデオクリップ320-322のシーケンスを示す。図8の横軸300は経過時間軸300であり、縦軸はデシベルで測定された音声レベルをしめしている (注: 図8は一定の縮尺で記載されていない)。ビデオカメラの良い機能は、映像とともに音も記録できることである。しかし、制作物を構成しようとすると、他の音声チャネルである音声バックキ

グチャネルを含まないものが普通である。背景音声を提供する追加バックトラックは、専門家が編集した結婚式のビデオのような、専門家が制作したビデオ制作物では当たり前の機能である。しかしながら、そういった編集は技術に通じたビデオ編集者が苦勞してまとめた上げた大変な作業である。

【0049】本発明の好適な実施形態において、ビデオクリップ320から322の間のトランジションは自動的に検出され、音声レベルも自動的に検出される。例えばビデオクリップ320に対応する音声レベル303はゼロデシベル(0dB)と検出され、ビデオクリップ321に対応する音声レベル304は無いと検出され、-10.0dBにセットされる。また、ビデオクリップ322に対応する音声レベル305は-3dBと検出される。

【0050】制作物ビデオを演出する際、あるクリップから次のクリップへのビデオトランジションに対応するそれぞれのポイント311において、予め定めた傾斜(ramping)310によって最終減衰波形313が生成される。最終減衰波形313は、好ましくはビデオ307のバックトラックとして使用される1つかそれ以上の楽譜のレベルセッティングの制御に用いられる。さらに、減衰波形はバックトラックがビデオに対応する音声から一定の時間残るよう、バックトラックに適用することもできる。さらに、もし望まれる場合には、バックトラックによって録音音声じゃまされないように、ユーザはビデオカメラで録音された音声と、バックトラックの音声レベルとの希望するレベル差を設定することができる。

【0051】<音声マッチング>好ましい実施形態において、第1の音声トラックは選択され、またこの選択された第1の音声トラックはまた、複数の音声トラックから予め定められた、実質的に第1の音声トラックと類似した音声トラックの検索及び選択に用いることができる。これら予め定められた音声トラックは、選択された第1の音声トラックと一緒に、或いはその替わりに用いることができる。複数の音声トラックから類似の音声トラックを決定する技術は、選択された第1音声トラックの特性と比較された各音声トラックの音声信号の特性に基づく。特性は、例えば平均ボリューム、ビート及び卓越振動数(dominant frequency)といったものの単独または組み合わせを含む。

【0052】以下、この音声マッチング規則がビデオ制作物にどの様に適用されるのかについて例を用いて説明する。例えば、製造者から選択された第1音声トラックとともに、ビデオ制作物のテンプレートが供給される。ユーザは、ユーザが好む複数の替わりの音声トラックを、テンプレートとともに供給された第1音声トラックに対してオファーすることができる。音声選択規則が第1トラックの1つかそれ以上の予め定めた特性に基づ

き、複数の代替音声トラックの中から最も適した音声トラックを選択する。

【0053】<映像コンテンツとモーション>ビデオにおいて、複数の重要な映像イベントは、そのビデオに任意に適用される所定の特殊効果において、ビデオのセグメントを特定する手がかりとされる。

【0054】例えば、ビデオセグメント(クリップ)と次のビデオセグメント間における好ましいトランジションのタイプをガイドするために、当分野で知られる動き検出技術がビデオに適用される。図9には、ビデオ500と、それに結合されたズームシーケンス501が各フレーム503の下に表示された相対ズーム比とともに示されている。

【0055】相対ズーム比502は、各フレームで検出された動きベクトルを用いて決定される。本発明の本質又は範囲を外れることなく、相対ズーム比502を例えば符号化されたビデオに付随したメタデータから計算することもできる。

【0056】符号化されたメタデータは、ビデオカメラ、具体的にはズームセンサ及び回路を、現在のズーム状況に対応するメタデータを供給可能なようにしたビデオカメラにより記録され、また得ることができる。

【0057】図10は、小屋601のビデオパンシーケンスからなるパンショット600を示す。ズームシーケンス501についての説明とほぼ同様の方法によって、パンショットは動きベクトルから検出することができる。ビデオカメラのパニングセンサがビデオカメラの動き情報を発生するようにしても良く、その情報をパンショットにおけるパン量の決定に用いても良い。

【0058】ビデオからのズーム及びパンショット検出において、相対ズーム比が所定フレーム数以内でズーム閾値を超えているズームのみを検出したり、相対パン移動量が所定のフレーム以内でパン閾値を超えるパンショットのみを検出したりすることを可能にするために、所定のズーム閾値及びパン閾値が任意に設定される。例えば、所定フレーム数を2、ズーム閾値を3とセットし、2フレーム以内で1:3を超える相対ズーム比が発生すれば、ズームシーケンス検出において検出されることになる。

【0059】パンショット又はズームが検出されたそれぞれのケースにおいて、ユーザに受け入れ又は拒否の選択を与えるようにし、それに応じて検出されたパン又はズームにマークをすることもできる。しかし、パンショット又はズームの位置を示すためにビデオにマークを付ける処理は、検出された各ズーム及びパンに所定の基準に基づいてマークするような自動化処理に代えることもできる。

【0060】<印刷によるビデオの要約>パンショット及び/又はズームのマークは、特殊効果を適用すべきビデオセグメントや、印刷のために印刷装置に読み出さ

れ、或いはビデオのアイコンを生成するために読み出されるビデオのセグメントを容易に識別するためのものである。例えば、図10のパンシーケンス600はつなぎ合わされ、すなわち、フレーム毎に合成され、図11に示される静止面パンを形成する。このビデオフレームのつなぎ合わせのプロセスは、当該技術分野において知られたもので、例えば、デジタイズされた表現において、ビデオの各フレームの所定の画像 (picture) 要素のトランジションを検出することにより行うことができる。

【0061】図12は、特徴的静止画 (salient still image) 704を得るための、ビデオのフレーム701～703の合成を示している。印刷装置で印刷するための特徴的静止画に合成されるフレーム701～703は、ビデオのズーム部分のフレームから抽出することが好ましい。図12に示される3つのフレーム701～703は、図9に示したズームシーケンスから選択され、特徴的静止画704を作るために用いられる。特徴的静止画704を合成するには、フレーム701で表される最初の画像を第1の所定ファクタで変倍し、変倍された画像は特徴的静止画704の第1部分705として用いられる。図12において示されるように、第1部分705は8×8の成分を指す。次に、フレーム702で表される第2の画像を第2の所定ファクタで変倍し、変倍された画像を特徴的静止画704の第2部分706となるよう、第1部分705に重畳或いは挿入される。最後に、フレーム703で表される第3の画像が第3部分707となるよう、第2部分706に重畳 (又は挿入) される。結果として得られる特徴的静止画704は異なる解像度から得られ、1つの特徴的静止画704を形成するために合成された画像704の3つの部分705～707から構成される。特徴的静止画704は印刷装置で印刷されても良い。

【0062】パンショット及びズームシーケンスが抽出されたビデオにおいて、そのビデオを撮影した写真家はカメラマンが特定のシーンをパン又はズームしたいと考えていたものと仮定すると、そのビデオにおけるパンショット及びズームシーケンスは特別な興味の対象である可能性がある。よって、図11及び12を参照して説明したように、それらの特別なシーンから得られた画像が印刷装置での印刷用に選択される。

【0063】全てのビデオがパンショット及び/又はズームシーケンスを含んでいるわけではないし、パン又はズームが特別に興味のある画像を含んでいないこともある。このような場合には、印刷装置で印刷しようとするビデオの画像は、動き解析や、フォーカス、スローモーションブラー (slow motion blur) といったビデオの他の特性に基づいて選択される。例えば、選択キーフレームにおいて起こる動きの局所最小値の決定は、カメラマンがシーンの中の興味ある場所で停止 (ポーズ) したり、重要性を強調するために人物がジェスチャーを休止した

りして選択キーフレームが興味の対象である場合の良い推定である。

【0064】<動きの選択されたトランジション>図13は、第1ビデオセグメント901及びその後第2ビデオセグメント902を表示する表示装置900を示す。最初に表示されたビデオセグメント901は表示装置900の右下角で発生する動きを有している。この動きは表示装置900の右下角に動きベクトルの第1セット903で示されている。表示装置900に表示された第2ビデオセグメントに対する動きベクトルの第2セット904は、第1及び第2のビデオセグメント901及び902の間で表示装置900の右下角から表示装置900の左側へ変化した動きを示している。

【0065】従って、第1及び第2のビデオセグメント901及び902が連続して表示される場合には、トランジションの発生が必要である。好ましい実施形態によれば、トランジション形式は表示されるビデオセグメントにおける動きの場所及び量によって選択される。第1ビデオセグメントと第2ビデオセグメントの間のトランジション形式は垂直又は水平及び/又は斜めのワイプ、フェード及び同様のトランジションを含む。

【0066】図14に、図13のビデオセグメントに対する水平ワイプトランジション905の例を示す。トランジション905において、第1ビデオセグメント901は去っていくシーンで、第2ビデオセグメント902は入ってくるシーンである。ビデオセグメント901及び902のそれぞれの動きの場所から、表示装置における右から左への水平ワイプトランジションが自動的に提案 (又は選択) される。右から左への水平ワイプトランジション905は視聴者の注意を第1ビデオセグメントで発生する動きから第2ビデオセグメントで発生する動きに効果的に移行させ、またトランジションが発生してゆっくりと表示装置900を横切っている間視聴者の注意を維持させる。トランジション形式はビデオセグメント間で起こる (目に見える) 動きの場所と量に基づき選ばれる。本発明の範囲及び本質を離れることなく、上述の例の替わりとなるいろいろな形式のトランジション及びトランジション選択に対応する動きの規則が考えられることはいうまでもない。例えば、視覚特性が大部分動きからなり、それぞれが中心的に配置されている2つのビデオセグメント (クリップ) を接続するような場合には、第2ビデオセグメントをフェードインしながら第1ビデオセグメントをフェードアウトするトランジションも可能である。

【0067】<カラーコンテンツ>好適な実施形態による、画像のコンテンツに基づく別の自動編集プロセスは色選択である。例えば、もしユーザがそうしたければ、ビデオ制作物のオープニングタイトルの色を、ビデオセグメントの導入部分もしくはビデオ全体に関する統計的カラーコンテンツ情報に基づいて選択することができる。

タイトルの色は、例えば補色の集合(complementary set of colors)又は補足色の集合(supplementary set of colors)から選択することができる。補色の集合又は補足色の集合はビデオから得られる複数の色として選択されることが好ましい。

【0068】図15には、画像1600と、その上に重ねられた“TITLE”という文字1601が示されている。画像1600は印刷画像、表示装置に表示された静止画又は処理されて表示装置に表示されたビデオの所定部分を表してよい。

【0069】図15にはまた、文字1601に対するカラーパレットを提供するのに用いられる、補色の集合1602および補足色の集合1603が示されている。補色1602及び又は補足色1603は画像1600のカラーコンテンツに基づき選択される。画像の優勢色を決定する既知の技術としては例えば、簡単なケースにおいては画像の所定領域に渡る色平均で優勢色を決定することができる。

【0070】図16には、画像のカラーコンテンツが興味の対象或いはそうでない、種々の部分401~405を有する表示装置400が示されている。例えば、もしグラフィカルな要素が文書のヘッダやフッタに類似した使われ方をするのであれば、401及び402で示される領域からカラーコンテンツ情報が抽出される。同様に、文字情報(或いはグラフィカルな要素)が画像の左の部分403又は右の部分404で表示されるのであれば、これらの対応するそれぞれの部分403及び404からカラーコンテンツ情報を抽出することが好ましい。

【0071】また、文字情報が中央に配置される場合は、画像の中央部405からカラーコンテンツ情報が抽出される。本発明の好適な実施形態において、カラーコンテンツ情報はまた、前述したビデオセグメント間のトランジションタイプの選択をガイドするためにも使用される。動き及び又はカラーコンテンツ情報はビデオセグメント間のトランジションタイプ選択に用いることができる。カラーコンテンツ情報をトランジションタイプの選択に用いる一つの例として、第1ビデオセグメントの平均カラーコンテンツ又はその代わりとしての輝度コンテンツが、続くビデオセグメントの平均カラーコンテンツと実質的に類似となるディゾルブトランジションを、第1ビデオセグメントとそれに続くビデオセグメント間のトランジションに選択する場合がある。

【0072】本発明の好適な実施形態はまた、図17に示されるコンピュータシステム1700の様な、通常の汎用コンピュータで実施されることが好ましい。この場合、図2~図6を用いて説明したプロセスは、コンピュータシステム1700で実行されるソフトウェアとして組み込まれる。コンピュータシステム1700はコンピュータモ

ジュール1701は少なくとも1つのプロセッサユニット1705と、一般的に半導体ランダムアクセスメモリ(RAM)及びリードオンリーメモリ(ROM)からなるメモリユニット1706と、ビデオインタフェース1707、キーボード及びマウス(又はジョイスティック)インタフェース1708を含む入力/出力(I/O)インタフェースを有する。

【0073】記憶装置1709は、フロッピーディスク、ハードディスクドライブ、CD-ROMドライブ又は同分野で知られた類似の不揮発性記憶装置の一つ又はそれ以上の装置を含むことができる。コンピュータ1701の構成要素1705から1709は、相互接続バス及び電算システム1700の通常の動作の結果として得られる、関係分野で知られた方法に乗取って通信を行う。実施形態を実行することのできるコンピュータの具体例としては、IBM-PC及びその互換機、サン社のスパークステーション又はそれらが進化した類似のコンピュータシステムが含まれる。

【0074】＜拡張収集装置＞図18は、拡張収集装置2100のブロック図である。拡張収集装置2100は、情報収集装置の一例としての構成として、コンテンツ収集部2101、処理部2104、メモリ記憶部2103、ユーザーインタフェース部2105、読み出し専用メモリ(ROM)2108、通信ポート2106及び出力インタフェース2102を有する。

【0075】拡張収集(キャプチャ)装置2100の所定の用途に応じて、通信ポートを除去し、記憶部2103をPCMCIA、バブルメモリ、フロッピーディスク又は任意の可搬型記憶装置に置き換えても良い。

【0076】生成(或いは収集)された出力情報は、出力インタフェース2102を介して外部装置へ伝達され、収集コンテンツを表すことができる。収集コンテンツ情報は外部装置の入力信号として使うことができる。例えば、外部装置としてはアナログ又はデジタルビデオレコード、光ディスクドライブ、音声録音装置、画像記録装置、プリンタ装置又は汎用コンピュータ(いくつか名前を挙げれば、IBM互換PC、アップル互換PC、サンワークステーション)といったコンテンツ記録装置を含む。収集コンテンツ情報はコンテンツデータを含み、好ましくはコンテンツデータに関連するメタデータをも含む。コンテンツデータの例としては、ビデオカメラ装置で収集された画像データであり、関連するメタデータとしては日付、時間、カメラセッティング及び画像を撮った場所などの情報を含むことができる。

【0077】コンテンツ収集部(CCU)2101は好ましくはそれぞれが所定形式のコンテンツ情報を受信(収集又は寄せ集める)ことが可能な、複数のコンテンツセンサを有する。このようなセンサの例としては、音声(コンテンツ)データ受信用のマイクロフォン、画像

データ収集用の電荷結合素子(CCD)、収集する画像

データに関連するライティングコンディション(メタデータ)を決定及び収集するための光センサを含む。CCU2101は更に種々のコンテンツセンサやビデオカメラの現在のズームファクタと言った関連するハードウェアの現状に関する知覚データを収集する要素(変換器)を有することが好ましい。この付加情報は処理部2104での使用に特に有効であり、収集コンテンツ情報のメタデータの一部としてエンコードすることができる。

【0078】ユーザはユーザインタフェース部2105を通して利用可能にしてある複数のオプションalテンプレートから少なくとも1つを選択する。ユーザは選択ボタン列(図示せず)を通して拡張収集装置2100をコントロールし、所望のテンプレートを選擇する。処理部2104はメモリ記憶部2103及び/又はROM2108と通信しながら、ユーザインタフェース部2105を介して拡張収集装置2100で現在有効な種々のオプションalテンプレートを提供する。

【0079】各テンプレートは一般に、所定のプログラム言語で記述されることが可能なテンプレートインストラクションの列から構成されている。例えば、そのプログラム言語があるアプリケーションに特有のもので、拡張収集装置2100が専有しても良いし、Java(サン・マイクロシステムズ社の登録商標)等の汎用プログラム言語によるインプリメンテーションであっても良い。テンプレートインストラクションは通信ポート2106を介して汎用コンピュータシステム2107から拡張収集装置2100へアップロードされ、メモリ記憶部2103に記憶される。

【0080】また、一般にコンピュータゲームに使用されているカートリッジシステム及び/又はフラッシュカードをテンプレート保存に使用し、拡張収集装置2100へテンプレートをアップロードするプラグインデバイスとして使用することもできる。好ましくは、テンプレートインストラクション列の一部、特にいくつかのテンプレートに共通なインストラクションを読み出し専用メモリ2108に記憶し、メモリ記憶部2103に必要な(RAM)メモリの容量を削減する要にすることが好ましい。このようにして、テンプレートインストラクションは処理部2104の実行時計算に使用することが可能になる。図18に示した汎用コンピュータシステム2107はコンテンツ情報をダウンロードする際に出力行タフェース2102とインタフェースする前述の外部装置としても使用されることが好ましい。

【0081】ユーザインタフェース部2105は例えば液晶ディスプレイ(LCD)及びコントロールボタンの列を有し、ユーザへ(メッセージ)情報を伝達し、同様にそれらのメッセージに対応するユーザ入力に使用される。

【0082】処理部2104はテンプレートインストラクションを解釈し、ユーザに有効なメニューの選択のた

めのいろいろな選択肢を表示及びユーザからのメニューへの応答を解釈しながら、コンテンツ情報の内容解析を含むオプションを提供する。

【0083】ユーザによって選択される選択肢は、CCU2101の現状、ユーザプリファレンス設定及び/又は拡張収集装置2100で使用されようとしている現在のテンプレートの項目に影響又は変更を及ぼす。これらタスクのいくつかは、テンプレートが有効なコンテンツ情報から最良の決定をし、それによってCCU2101を修正するように完全に自動化されることが好ましい。例えば、以下に説明する拡張収集装置の第1実施形態において、所定の状況に基づいて、ランドスケープ表示の画像の上にポートレート表示の画像を配置が希望されているので、必要ならポートレートの画像から端を切り取るという判断にテンプレートを用いても良い。

【0084】図28は収集情報の拡張操作全体を示すフローチャートである。プロセスはステップ21100で開始する。ステップ21102において、少なくとも1つのテンプレートが拡張収集装置にアップロードされる。ステップ21104において、拡張収集装置において少なくとも1つの形式のコンテンツ情報が受信される。ステップ21106ではアップロードされたテンプレートのテンプレートインストラクションが実行され、受信したコンテンツ情報に基づくインストラクション構成物が生成される。ステップ21108ではコンテンツ情報の構成要素がインストラクション構成物に従って拡張される。そしてステップ21110で処理が終了する。拡張されたコンテンツ情報を記憶手段への記憶に関するを更に含むことが好ましく、また記憶されたコンテンツ情報が外部装置へダウンロードされるようにしても良い。

【0085】<第1の拡張収集装置>以下、拡張キャプチャ(収集)装置2100が内蔵されるか、拡張システムの一部を形成しているデジタルスチルカメラを例にして、第1の拡張収集装置について説明する。本実施形態においてコンテンツ収集部2101はこの種のカメラの一般的構成と同じく光学レンズシステム及びCCDを有し、この他に測距センサ、光強度測定装置、フォーカス検出装置及び位置検出装置(例えばグローバルポジショニングシステム:GPS)を含むことができる。これらのセンサ及び装置はイメージデータ及び撮影場所などのメタデータを収集する。デジタルスチルカメラのユーザインタフェース部2105は液晶ディスプレイ(LCD)であり、好ましくはユーザインタフェース部2105はデータ入力装置(例えば、ボタン)を含む。通信ポート2106は1つかそれ以上の必要なテンプレートを汎用コンピュータ2107からアップロードし、メモリ記憶部2103に記憶するために用いられる。

【0086】図19はデジタルスチルカメラによって被写体2201の写真を撮っている状態を示す模式図であ



る。デジタルスチルカメラ2202はその一部に、記録される画像が表示されるLCD2203を有する。ユーザは記録する画像をシャッターボタン2206の押下によって選択するか、モードテンプレートボタン2204の押下によってテンプレートモードの入力を行うことができる。ユーザは2度目にモードテンプレートボタン2204を押下することによってテンプレートモードから出ることができる。

【0087】図20に示すように、テンプレートボタン2204を押下するとLCD2203に使用可能なテンプレート2301のリストが表示される。本実施形態においては、テンプレートはランドスケープテンプレート、パンショットテンプレート及びポートレートテンプレートを含み、前述のように前もってアップロードされている。セレクトボタン2205は、有効な選択肢の間をナビゲートし、希望のテンプレートを選択するのに使用される。LCD2203に表示された選択アロー2302は現在のテンプレートモード選択を示す。

【0088】テンプレートの選択、たとえばポートレートテンプレートの選択がなされると、LCDに表示された画像が処理部2104で解析される。解析結果はテンプレートがユーザをガイドするのに使用する。本実施形態において、被写体2201が中心に位置し、かつ表示装置の可視領域（すなわち、記録される画像）の少なくとも所定パーセンテージを占めるような最良の結果を得るためにはポートレートテンプレートが必要である。

【0089】図21はポートレートテンプレートが選択された後のLCD2203を示す図である。ポジショニングインジケータ2401、ラインアップマーカ2404及びセンターマーカ2405を含むいろいろな要素が液晶表示装置2203に現れる。これらの要素はユーザがシャッターボタン2206を押すことによって画像の記録を取ろうとした場合に、写真（または記録された画像）の上に現れない。ポジショニングインジケータ2401、ラインアップマーカ2404およびセンターマーカ2405の現況の詳細が生成され、後の使用または評価のために、対応する画像に関連するメタデータとして含まれることが好ましい。例えば、ユーザに関連するマーカ2404、2405及び/またはポジショニングインジケータ2401を正しく整列しておらず、コンテンツ収集ユニットが希望するよりも大きなビューの画像収集が可能である場合、マーカ2404、2405及びポジショニングインジケータ2401の情報に関する、メタデータの情報部分を画像のより好ましい印刷範囲の評価に用いることができる。そして所定の印刷範囲に合わせて画像の端を切り落としてもよい。

【0090】好ましくは、画像及び対応するメタデータは、必要なら所定の印刷範囲に切り揃えて出力装置（例えば印刷装置）に出力する前にまず画像編集に使用される汎用コンピュータにダウンロードされる。

【0091】ポジショニングインジケータ2401はインジケータ上のペリフェラルアロー2402を点滅させてユーザに異なる方位、例えばカメラ2202の垂直方位の使用を、アドバイスしてもよい。ポジショニングインジケータはその他にもズームのアドバイスをユーザに行う。ズームインアロー2403は画像の被写体2201のズームインが望ましいことを示す。従って、ズームアウトが望ましいことを示すためにズームアウトアロー2406を点滅させてもよい。その代わりに、被写体2201がカメラ2202と共に近づけるか遠ざけて、ズームインやズームアウトとそれぞれ実質的に同等の効果を達成してもよい。

【0092】センターマーカ2405及びラインアップマーカ2404は、画像のコンテンツ情報の分析及び画像上の被写体2201のおおよその位置を実質的に検出することが可能な拡張キャプチャ装置2100を有する第1の拡張収集装置のバリエーションを提供するために設けられている。

【0093】画像の被写体2201の大きな位置検出にコンテンツ情報の解析を行う必要はなく、他の方法によって行うことができる。例えば、ユーザが画像のどの部分を見ているのかを、観察される画像に関するユーザの瞳孔の相対位置から検出し、画像中の被写体位置決定を行うことができる。

【0094】第1の拡張収集装置のこのバリエーションにおいて、センターマーカ2405は可視領域の中心に、ラインアップマーカ2404は検出された被写体の中心に配置される。ラインアップマーカは被写体に追従しようとするので、ユーザはカメラを動かすことによりラインアップマーカ2404とセンターマーカ2405が並ぶように操作することができる。

【0095】図22に示されるように、ユーザはペリフェラルアロー2402の点滅が止み、正しい方向を表すようにカメラ2202を希望する垂直方向に回すことが好ましい。この段階では、図22に示すように、マーカ2404と2405を整列させる必要はなく、被写体2201は画像の所定部分を占めておらず、またズームインアロー2403が依然としてズームインが望まれることを示している。

【0096】図23は、ユーザが正しく画像の被写体2201を中心に配置し、ズームインを行ない、ポジショニングインジケータの全てのアローが消え、好ましくは整列のインジケータ2602がLCD2203に表示された結果を示す。ズームや方向付けに関する他のいかなる望ましい変更も表示しないが、ポジショニングインジケータは選択されたテンプレートに従って画像（写真）の構成要素がズーム及び方位について正しいのか、また現在の画像（写真）が記録に推奨されるかについての示唆に富む。そして、ユーザはシャッターボタン2206をアクティベートして画像を記録する。

【0097】図24は、画像の放写体が1つより多い要素から構成されている場合の、ポートレートテンプレートモードにおけるデジタルスチルカメラ操作例を示す。テンプレートは画像の視覚領域に対する構成物の最良なはめ込みを最適化しようとする。好ましくは、構成要素を含む領域ボックスの幾何学的中心は整列マーカ2702の場所として使用され、また画像の可視領域（すなわち、可視表示領域2203）の幾何学的中心はセンターマーカ2701として使用される。さらに、構成要素の一つが部分的に可視表示領域2203の外に出ているので、ポジショニングデータがズームアウトアロー2406を点灯させてズームアウトが望ましいことを表示する。

【0098】実質的に類似の説明が、一般に風景写真に使用されるランドスケープテンプレート及びパノラマ写真に用いられるパンショットテンプレートについても適用される。簡潔で明快な説明のために、テンプレートはオリエンテーション、ズーム等、少ない数の機能を提供するものとして説明されてきたが、テンプレートはユーザにアドバイスをを行うこともできる。しかしながら、テンプレートはこれらに限定される物ではない。例えば、1つかそれ以上のテンプレートは、写真に対する望ましいライティング、写真スタジオタイプ設定におけるライティングの最良アングル及び写真の構図を向上するためのカメラフォーカスのアドバイスをを含む機能についてユーザにアドバイスを提供するようにしても良い。カメラフォーカスのアドバイスは、バックグラウンドフォーカスとフォアグラウンドフォーカスを達成するカメラセッティングを提案する方法で行うことができる。

【0099】テンプレートが有効なコンテンツ情報からユーザに与えようとするアドバイスは、当業者が通常なし得るアドバイスであってもよい。例えば、一般に高品質の写真や予め定めた形式の写真を制作することに不慣れたアマチュア写真家へのガイダンスを提供するためのアドバイスであって、好ましい方位付け、最良のズーム及び被写体の配置のアドバイスなどを挙げることもできる。

【0100】前述したとおり、テンプレートはプログラミング言語を使用してコードにインプリメントされる。後述する添付Aにポートレートテンプレートの疑似コードが含まれている。疑似コードのメインルーチンはデジタルスチルカメラの起動中に動作し、有効なテンプレート機能が最適かどうかを検出するために使用される種々のコンディションをチェックするため、ビューファインダー（LCD2203）の画像を解析する。

【0101】ユーザがテンプレートモード選択ボタン2204を押すと、有効なテンプレートが表示される。ユーザは希望するテンプレートを選択ボタン2205により選択することができ、そしてテンプレート選択モードを終了する。選択を行う際、補足Aに含まれる疑似コ

ドで表されるようなソフトウェアルーチンがビューファインダーの画像のズーム及びオリエンテーションに必要な種々のコンディションのチェックを行う。

【0102】SetUpForTemplateサブルーチンは画像が整列されているときに真の論理値を返し、メインプログラムのVoidMain()が画像（ピクチャ）が整列されているというフラグを立てる。

【0103】前述したように、メタデータのような付加情報を記録し、画像に対応付けすることができる。この付加情報の例としては、日付、時間、使用されたズーム、シャッタースピード、絞り、静止画が人物のポートレートであるか、画像の中の人物の数、検出された全てのそれぞれの顔の中心はもちろん、検出された全ての顔の幾何学中心をも含む。

【0104】＜第2の拡張収集装置＞以下に述べるように、第2の拡張収集装置は拡張収集（キャプチャ）システム2104がビデオカメラの一部を構成することを除き、第1の拡張収集装置と同一である。テンプレートはビデオクリップのズーム、パン、合成及び演出（映画演出に関して）のアドバイスを提供できるように供給される。例えば、ビデオカメラへのアップロードのための、複数のアップロード可能なテンプレートをユーザに提供することができる。このテンプレートには結婚式シーンの演出用、不動産業者の住宅販売プロモーションクリップ用、田舎の休日風景用（ビデオに録画することが推薦される世界の風景や名所の提案を含んでもよい）テンプレートを含む。

【0105】具体的には、ユーザが休暇にエジプトへ拡張収集システム2104を有するビデオカメラを持って出かけてもよい。ユーザがエジプトにいる間に撮影されるビデオ録画の要素を拡張するため、北アフリカの好ましい風景のテンプレートを拡張収集装置2104にアップロードすることができる。ここで説明される他の機能の実行以外に、テンプレートが更に、録画する価値において推薦される風景や名所（例えば、ピラミッド）の提案を行ってもよい。

【0106】更に別の例として、結婚式のシーンを演出するためのアップロード可能なテンプレート（以下、ウェディングテンプレートとする）のために作成されたリファレンスがある。親戚の結婚式のビデオ記録を作りたいと希望するユーザは、汎用コンピュータにウェディングテンプレートと呼び出して、それを拡張収集システム2104を有するビデオカメラに通信ポート2106を介してアップロードする。ウェディングテンプレートは（コンパイルされた）コードとしてメモリ記憶手段2103に記憶される。一方、アップロード可能なテンプレートに対する一般的な機能は前もって、例えばビデオカメラ拡張収集装置の製造者によってROM2108に記憶されている。一般機能には、有効な最小及び最大ズーム設定、可能な絞り設定及び大半のビデオカメラ用テ

ンプレートに共通と思われる他の機能を含む。

【0107】ユーザがビデオレコーダーを用いてウエディングシーンを記録し始めると、ウエディングテンプレートが起動され、専門家のビデオ記録に応じたビデオクリップを制作するために録画しなくてはならない人のショットについてのアドバイスをユーザに与える。例えば、ユーザは希望するウエディングシーンの要素を提供するショットとともに、好ましくは時系列的に、注意を促される。すなわち、花婿が教会に入ってくるショット、花婿が祭壇に立っているショット、花嫁が車から出て教会の階段を進むショット、花嫁が教会の通路を歩くショットなどを撮影するように促される。

【0108】さらに、ユーザはズーム、パニング、ライティング及び音声といったビデオ技術に関してもアドバイスされる。例えば、花嫁が教会の通路を歩いているショットを撮影している時、もしユーザが被写体を早くズームしすぎていれば、ビデオカメラのビューファインダーに“速度を落とせ”というインジケータが点灯し、逆にズームが遅すぎればズームのスピードを上げるよう同様に知らされる。

【0109】好ましくは、ビューファインダー(LCD)に表示可能なメニュー駆動型コマンド列を通して、今撮影している場所や部分をユーザが指示する。ユーザはこれから撮影する、或いは今まで撮影していた部分に印を付ける(選択する)。例えば、ウエディングテンプレートにおいて、ユーザには花嫁が通路を歩いているとき、通路部分を含む、撮影することが推奨される多くの場所や部分の表示が与えられ、ユーザは関連するシーンの撮影前或いは撮影後に通路部分に印を付ける。場所に印を付けたことに関連したデータはメタデータとしてビデオクリップの対応する部分に記録される。例えば、花嫁が教会の通路を歩いている場所であるウエディングシーンの部分(場所)という情報は、メタデータの一部としてウエディングシーンに記録される。録画されたウエディングシーンの編集、好ましくはビデオクリップ編集に適応されたコンピュータのような外部装置による編集において、ウエディングシーンに関連したメタデータはウエディングシーンのいろいろな部分を見分けるために使用される。

【0110】<第3の拡張収集装置>第3の拡張収集装置は拡張収集装置を有する音声ミキシングデスクを参照して説明する。音声ミキシングデスクは音楽コンテンツを複数のチャンネルから収集し、個々のチャンネル及び/又はチャンネルグループに対して様々な効果やフィルタを適用し、これらのチャンネルを1つかそれ以上の出力チャンネルに結合するのに一般的に使用される。出力チャンネルはその後聴衆のために録音され、またはアンプの駆動に用いられる。

【0111】簡単のためにもう一度、単純な2チャンネルインターフェース音声ミキシングデスクを説明する。

ミキシングデスクは、コントロールの表示、フィルタ、アナライザ、エフェクトユニット及び類似の補助ハードウェアを用いたコンテンツ解析、ユーザフィードバックの提供及びフィルタ及びエフェクトユニットへのコントロール信号の送信のためのコンピュータによって実現される。

【0112】図25は、音声ミキシングデスクシステム2800を示す。音声ミキシングデスクシステム2800は、音声変換器の様な複数のコンテンツ収集センサ2801と、複数チャンネルをサポートできる拡張収集装置2802と、汎用コンピュータシステム2804と、複数のエフェクトユニット、フィルタバンク及びアナライザ2810と、少なくとも1つの出力ポート2811を有する出力増幅器2809と、コンパクトディスクCDプレーヤ、オープンリール及び/又はカセット形式のテープユニットを含む再生装置2812を有している。

【0113】センサ2801及び他の周辺機器2812で収集された情報は、選択されたテンプレートに従って収集情報を処理する拡張収集装置2802へ送られる。所望のテンプレートコレクションは(総括的な)データリンク2805によって、拡張収集装置2802のヘアップロードされ、装置2802のメモリ記憶部に記憶される。コンテンツ収集センサの形式は実施毎に変わり得るが、拡張収集装置2802は図18と実質的に同じ構成を有する。本実施形態においては、図18におけるユーザインタフェース部を取り除き、替わりに外部表示装置2813、キーボード2814及びマウス装置2816を設けることが好ましい。

【0114】センサ2801及び他の周辺機器2812で収集された情報はコンテンツ情報の一部を形成する。コンテンツ情報は好ましくは時間、日付、録音期間の長さ、音楽のスタイル及びトラックのタイトルといったメタデータを含む。これらのメタデータは選択されたテンプレートによってコンテンツ情報に追加されても良い。複数のテンプレートが拡張収集装置2802へアップロードされるが、表示装置2813に表示されるマウス及び/又はキーボード装置駆動型のメニューの操作によってユーザが選択できるテンプレートは1つである。収集された情報は拡張収集装置2802と複数のエフェクトユニット、フィルタバンク及びアナライザ2810との間の通信リンク2803を介して出力増幅器2809及び出力ポート2811へ送られ、使用される。コンテンツ情報は総括的リンク2805を通してコンピュータ2804へ送られ、コンテンツ情報に関するデータがユーザのために表示装置2813上に表示される。コンテンツ情報は更に表示すべきデータを抽出する中央処理装置で解析されることが好ましい。あるいは、コンテンツ情報は拡張収集装置で解析され、関連する表示データのみがリンク2805を介してコンピュータ2804へ伝送される。

【0115】ユーザは入力装置2806及び2814を操作して様々な設定をその場で変更することができる。種々の設定は、両方向リンク2807を介して、2811の情報出力への所望の変更を生成するようににエフェクト及びフィルタのパラメータを変更するようにに設計された出力増幅器2809及び複数のエフェクトユニット、フィルタバンク及びアナライザ2810へコントロール信号として伝送される。双方向リンク2807は複数のエフェクトユニット、フィルタバンク、アナライザ2810、出力増幅器2809及び再生装置2812を含む多数の装置が生成した信号及び、必要なら様々な再生装置2812の現在のステータス及び/又は再生コンテンツ情報信号をも伝送することが好ましい。

【0116】ユーザはプリファレンス及びデフォルト設定をコンピュータシステム2804を通して変更することができる。デフォルト設定は一般に音声ミキシングデスクに接続された構成要素の種類及び/又は組み合わせによって規定される。具体的には、変換器又は再生装置の異なる選択が異なるデフォルト設定に必要とされる。プリファレンス設定は一般に音声ミキシングシステムをユーザの好みの設定にカスタマイズする方法を提供する。プリファレンス及びデフォルト設定は通常コンピュータシステム2804の一部である不揮発性メモリに記憶される。

【0117】バンドミックステンプレートを例として、アップロード可能なテンプレートの例を以下に説明する。バンドミックステンプレートは音声ミキシングシステムのユーザに対し、ストリングカルテット、ロックンロールバンド、ヘビーメタルバンド及びクラシックオーケストラ等のバンドスタイルを含む音楽バンドの音声ミキシング（希望があれば録音についても）に関連したアドバイスを提供する。バンドミックステンプレートの選択（及び実行）において、ユーザはバンドの音楽スタイル、場所の大きさ（すなわち、小ホール、野外、スタジオ等）の印付け（選択）を促される。好ましくはユーザは更にどのチャンネル（又は対応するセンサ）が楽器の情報（音声信号）の収集に使用されるか及びどのセンサがボーカル用であるのかの指定を促される。そして、バンドミックステンプレートは各チャンネルで使用可能なマイクロフォン（センサ）の最良な配置を提供する推奨リストを生成する。

【0118】さらに、ユーザはテストラン（ドライラン）を選択することができる。テストランにおいて、バンドミックステンプレートはコンテンツ情報を解析してどのセンサが離れた周囲の物体の反響を拾っているかのアドバイス及び解決方法（例えば、マイクロフォンの再割り当て又はセンサの再配置）の推薦もしくは使用可能なエフェクト及びフィルタによる歪みの電氣的な補償を試みることもできる。同様に、バンドがドライランを正しい順序で演奏した場合には、音声レベル設定が決定で

き、結果として得られるインソウのガイドとして使用することができる。

【0119】センサ2801で収集されたデータ（情報）について、拡張収集装置2802はコンテンツ情報の抽出を行い、このコンテンツ情報をコンピュータ2804へ供給する。好ましくは、得られたコンテンツ情報からコンピュータ2804が各チャンネルのセンサに対する、フィルタ（グラフィックイコライゼーション）の調整を行うための補償制御信号を各チャンネルに送信する。コンピュータ2804はまた、テンプレートの音楽スタイルに適合するように、エフェクトを調整しかつ各チャンネルのフィルタをも微調整する制御信号を、バンドの全ての楽器へ送信する。

【0120】例えば、ユーザはボーカルチャンネルに高ディストーションレベルを有するヘビーメタル音楽テンプレートを選択することができる。トップエンド周波数域の増強（ブースト）がミキシングデスクのユーザに推奨され、好ましくはこれらの機能はユーザが別の物を希望しない限り自動的に採用される。

【0121】しかしながら、もしユーザがクラシック音楽テンプレートを選択した場合には、全てのチャンネルは音声レベル設定ができる限り平坦になるようにバイアスされる。すなわち、楽器の原音にできるだけ近い音が再生されるよう、エフェクトユニット及び/又はフィルタバンクを通じた一切の（特殊な）エフェクトをチャンネルに加えない。

【0122】前述したように、テンプレートは技術分野の当業者が通常行うようなアドバイスをユーザに提供する。本実施形態において、技術分野は音声ミキシングの分野であり、例えばエコー、反響、残響、テンポに関するアドバイスは少なくともサウンドエンジニアが録音及びミキシングの際に行うであろうアドバイスの少なくともいくつかの要点である。

【0123】第2の拡張収集装置に関連して若干説明したに過ぎないが、上述の各装置に適用可能なテンプレートの好ましい任意付加機能としては、予め定めたイベント又はアクションに関してユーザが印付け（選択）可能なイベント又はアクションのチェックリストの提供がある。具体的には、上述のデジタルスチルカメラの例において、予め定めたショットのチェックリストがユーザインタフェースを介して供給される。ユーザがショット（写真）を撮る前（又は後）、ユーザはチェックリストから対応する所定のショットにチェックマークを付ける。

【0124】例えば、デジタルスチルカメラには誕生パーティーチェックリストを含む誕生パーティーテンプレートがアップロードされているとする。チェックリストはユーザインタフェース上に表示され、例えばチェックリストは以下の様な項目を含むことができる。

・誕生パーティーゲームショット

ろうそくを吹き消すショット  
 ・誕生日を迎えた人のために乾杯をしているショット  
 【0125】ユーザが誕生パーティの写真を撮り、もし撮影した写真がチェックリストに記載されたショットと一致した場合には、ユーザはチェックリストに記載され、撮影が済んだ所定の写真にチェックマークを付ける（又は選択する）ことができる。チェックリストから選択したショットにチェックマークを付けることによって、メタデータが生成され、写真（又は画像）に関連づけられる。このメタデータ情報は例えば誕生パーティのアルバムを構成するために写真（又は画像）が汎用コンピュータにダウンロードされる場合に利用することができる。

【0126】写真のアルバムを構成する際、どの写真が良い構成を形成するかの決定においてユーザをアシストすることができるテンプレートのそれぞれについて、複数のチェックリストを提供することができる。

【0127】サウンドミキシングデスクの例において、音楽に対する実質的に類似のチェックリストを提供することができる。すなわち、ユーザがリストに記載されたイベント（例えば、音楽的な断片（musical piece）やミキシング特性）が演奏される毎にチェックリストにチェックマークを付ける。さらに、対応するメタデータが生成され、イベントに関連づけられる。対応するメタデータは、イベントの単純な識別子でも、後で行われるイベントの操作や編集に望まれるであろうイベントのもっと複雑な記述であってもよい。

【0128】図26は、ユーザインタフェースディスプレイ（例えば図19に示したデジタルスチルカメラのビューファインダ2203や、前述した第2の拡張収集装置のビデオカメラのビューファインダ）に表示されたチェックリストの別の例を示す。チェックリスト2900は第1及び／又は第2の拡張収集装置の説明において記述した拡張収集装置2100にアップロード（ロード）されることができる不動産テンプレートに関連付けされたものである。不動産テンプレートはデジタルスチルカメラ又はビデオカメラのユーザ（一般に不動産販売員）に、合成及び／又は良いプレゼンテーションを構成するのに必要な画像についてアドバイスする。不動産テンプレートは任意の付加機能として、記録が推奨される写真（又はビデオカメラの場合はビデオクリップ）の簡単な説明を提供する説明フィールド2901、対応する記録された画像（写真）のそれぞれをリストから消すためのチェックリストインジケータフィールド2902及び既に記録された推奨される画像を示すためのチェックリストマーク2903を含む不動産チェックリスト2900を提供することができる。

【0129】チェックリストマーク2903は好ましくは推奨された画像が記録されたときに、有効なユーザ入力インタフェース（例えば、ボタン）を通してユーザが

加えるのが一般的である。不動産テンプレートはチェックリストの項目が消されると、メタデータを生成してチェックリストの対応する画像に関連づける。メタデータは関連づけられた対応画像の符号化された説明であることが好ましい。さらに、メタデータは時間、日付及び場所といった特性を含むことができる。

【0130】図27は、拡張収集装置21000から、アップロード可能なテンプレート21002を有するコンピュータシステム21001及び出力装置21003に続く情報の流れを示すブロック図である。複数のテンプレート21002が双方向通信リンク21005を介して、一般にテンプレートのセット21002を有するコンピュータシステム21001からアップロードされる。

【0131】コンテンツ情報21006は入力装置21000で収集され、メタデータ21007は少なくとも1つのアップロードされたテンプレート21004によって生成される。生成されたメタデータ21007は、チェックリストへの応答によって決定されるコンテンツ情報及び／又はコンテンツ情報の分析によってテンプレートが決定するコンテンツ情報の説明と同じくらい単純でよい。生成されたメタデータ21007はテンプレートによってコンテンツ情報の対応する部分又は全てに関連づけられる。例えば、不動産テンプレートの場合、そのコンテンツ情報に不動産テンプレートが提案する画像の推奨リスト2900からの画像を有するであろう。画像に対して生成される対応するメタデータは、オリエンテーション、平均色、平均輝度、支配的な色、所定のパターン、空間の統計学的分析、コンテンツ情報で区別される時間又は周波数ドメイン、色を含むユーザプリファレンス、情報の配置及びハードウェア要素の現在のステータス、様々な外部装置にダウンロードするための情報の所定フォーマットを含む装置固有の特徴を含む要点に基づいて作成することができる。

【0132】コンテンツ情報及びメタデータを含む拡張収集情報は外部装置にダウンロードされる。本実施形態において、外部装置は先述したように汎用コンピュータシステム21001である。

【0133】しかしながら、外部装置は拡張収集情報の受信専用の外部装置であってもよい。このような専用装置は、それぞれが収集情報の受信、所定のプリファレンスに従った情報の出力又は記憶の提供に適合された、プリンタ装置、表示装置（例えば、陰極管（CRT）、液晶表示装置（LCD））、電子複写装置及び／又は音響記録装置を含む。外部装置としては例えば最終的な出力を構成する後段において使用される拡張収集情報の記録及び記憶のためのビデオカセットテープでもよい。

【0134】アップロードされたテンプレート21004は拡張収集情報を所定の設定に従ってフォーマットかつ配列するか、拡張収集情報を拡張収集情報がダウンロ

ードされる外部装置の種類に応じてフォーマットすることが好ましい。

【0135】拡張収集情報の処理はマイクロプロセッサのような処理装置によって読み込み及び実行が可能な、コンピュータが読むことのできる記録媒体に記録されたソフトウェアで実施されてもよい。記録媒体としては、ハードディスク及び/又はフロッピーディスクドライブ、磁気テープ、CD-ROM、光磁気ディスク、メモリカード、読み出し専用メモリ等の磁気及び光学記憶メモリを含む。更に、ソフトウェアはネットワークシステム又はインターネットを通じてアクセス可能な別のコンピュータシステムにあっても良い。

【0136】本発明の実施に使用可能な拡張収集装置の例としては、以下のデジタルビデオカムコーダ装置を含む。

ソニー VX-1000 (ソニー株式会社製)

ソニー PC-1Q (ソニー株式会社製)

パナソニック NV-DS5 (松下電器産業株式会社製)

\* キヤノン MV1 (キヤノン株式会社製)

キヤノン XL1 (キヤノン株式会社製)

【0137】これらの装置は記録された音声及び画像情報に関連したメタデータの限られたセットの生成を提供する。そのようなメタデータは一般に情報の基本的な識別に関連したものである。更に進んだ装置としては、

【0138】

コダック DC220, 260 (米コダック社製)

ミノルタ Dimage EX Zoom 1500 (ミノルタ株式会社製)

を含む。これらの装置は静止画デジタルカメラであり高レベルのメタデータ、例えば図26を用いて説明した不動産に関する実施形態に対応したメタデータを生成するであろう。

【0139】以上の説明は本発明のいくつかの実施形態を説明したに過ぎないが、それらを修正及び/又は変更したものは本発明の範囲及び精神に含まれる。

【0140】

```

*
  付録A
  疑似コード
  // ポートレートテンプレート
  Boolean SetUpForPortrait(Boolean UI_Flags[], Image)
  {
    Boolean Result = TRUE
    // Assume only called IF 0 < numFacesDetected < 4
    IF InadequateMarginAroundGeometricCentre(image) THEN
      UI_Flags(ZOOM_OUT) = TRUE
      Result = FALSE
    ELSEIF TooMuchMarginAroundGeometricCentre(image) THEN
      UI_Flags(ZOOM_IN) = TRUE
      Result = FALSE
    ENDIF
    IF ((FacesBoundingBoxTallerThanItIsWide(image) AND
      CameraOrientation Equals LANDSCAPE) OR
      (FacesBoundingBoxWidererThanItIsTall(image) AND
      CameraOrientation Equals PORTRAIT)) Then
      UI_Flags(CHANGE_ORIENTATION) = TRUE
      Result = FALSE
    ENDIF
    IF (FacesBoundingBoxCentred(image) EQUALS FALSE THEN
      UI_FLAGS(CENTRE_BOUNDINGBOX) = TRUE
      Result = FALSE
    ENDIF
    RETURN Result
  }
  Void main()
  Boolean UIFlags(CONST_NUM_UI_FLAGS)
  ImageType image
  TemplateType templateSelected

```

```

3:9
WHILE PowerOn
  IF MenuButtonPressed() Then
    templateSelected = DisplayAvailableTemplates()
  ENDIF
  // カメラが見ている画像の取得
  GetImageInViewfinder(image)
  // テンプレートがある場合のチェック
  IF templateSelected EQUALS CONST_PORTRAIT_TEMPLATE THEN
    IF SetUpForPortrait(UIFlags, image) EQUALS TRUE THEN
      UIFlags(PICTURE_ALIGNED) = TRUE
    ENDIF
  ELSEIF .....
    ..... Other checks for other templates
  ENDIF
  UpdateUI(UIFlags)
END WHILE

```

【図面の簡単な説明】

【図1】 本発明の好ましい実施形態に係る装置の概要を示すブロック図。

【図2】 好ましい実施形態に従って生成されたマルチメディア要素の入力スクリプトのサンプルを示す図。

【図3】 好ましい実施形態におけるビート同期処理のフローチャート。

【図4】 好ましい実施形態におけるトランジション生成処理のフローチャート。

【図5】 好ましい実施形態における編集処理の概要を示すフローチャート。

【図6】 図5の編集処理の一部をより詳細に説明するフローチャート。

【図7】 好ましい実施形態におけるマルチメディア要素間におけるビート同期処理の例を示す図。

【図8】 好ましい実施形態における音声レベル編集例を示す図。

【図9】 画像要素の一部を構成するズームシーケンスの例を示す図。

【図10】 画像要素の一部を構成するパンシーケンスの例を示す図。

【図11】 図10のパンシーケンスをつぎあわせから得られる画像を示す図。

【図12】 図9におけるズームシーケンスの要素から得られた代表的静止画を示す図。

【図13】 表示装置に表示されたときの動きを示す2つのビデオシーケンスの動きベクトルを示す図。

【図14】 ビデオシーケンス（セグメント）間のトランジション例を示す図。

\* 【図15】 好ましい実施形態に従って選択された色を有するオーバーレイ画像"TITLE"と、画像の例を示す図。

【図16】 画像から色コンテンツ情報を抽出するための適当な対象領域の例を示す図。

【図17】 好ましい実施形態の実行に使用されるコンピュータシステムを示すブロック図。

【図18】 拡張収集装置の概要を示すブロック図。

【図19】 図18の拡張収集装置を有するデジタルスチルカメラを示す図。

【図20】 図19のデジタルスチルカメラのテンプレートセクションモードを示す図。

30. 【図21】 図19のデジタルスチルカメラの特徴を構成するオリエンテーションガイダンス及び整列を示す図。

【図22】 図19のデジタルカメラの他の特徴を構成するズームガイダンスを示す図。

【図23】 拡張画像構成の例を示す図。

【図24】 画像に一人より多い人間がいる場合のポートレートにおけるビューファインダの表示例を示す図。

【図25】 他の実施形態に係る拡張収集装置を示すブロック図。

40. 【図26】 テンプレートに関連づけられたチェックリストの例を示す図。

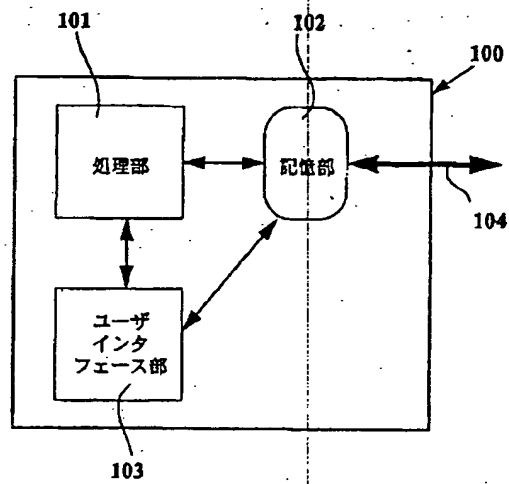
【図27】 拡張収集装置における情報の流れを示すブロック図。

【図28】 収集された情報の拡張処理を示すフローチャート。

(22)

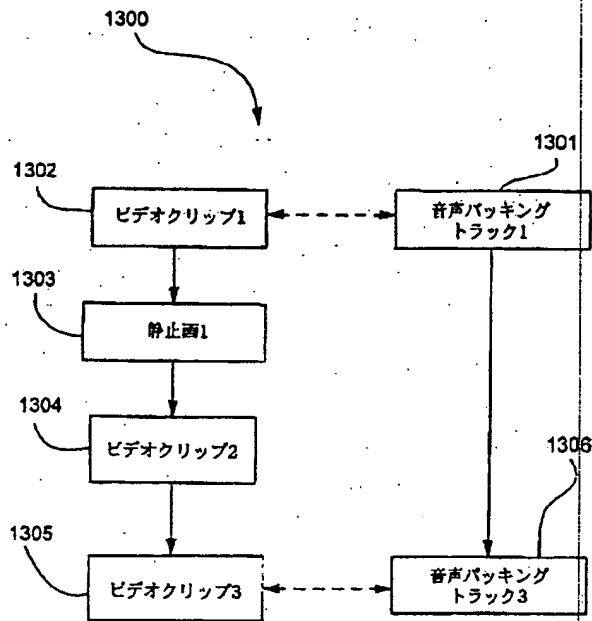
特開平11-234560

【図1】

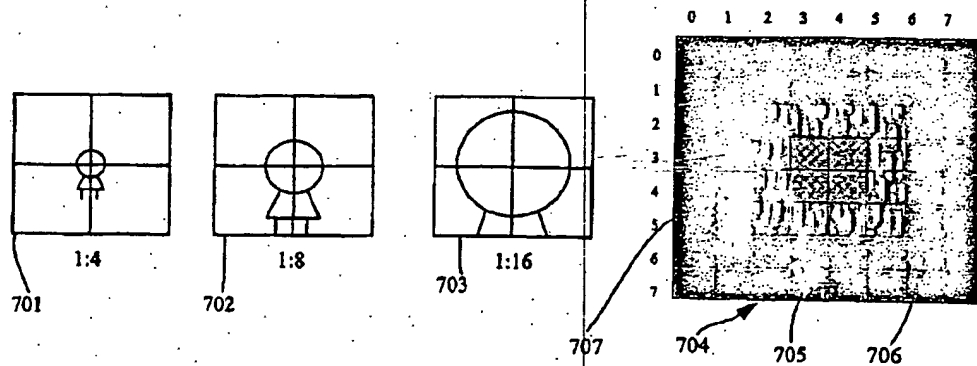




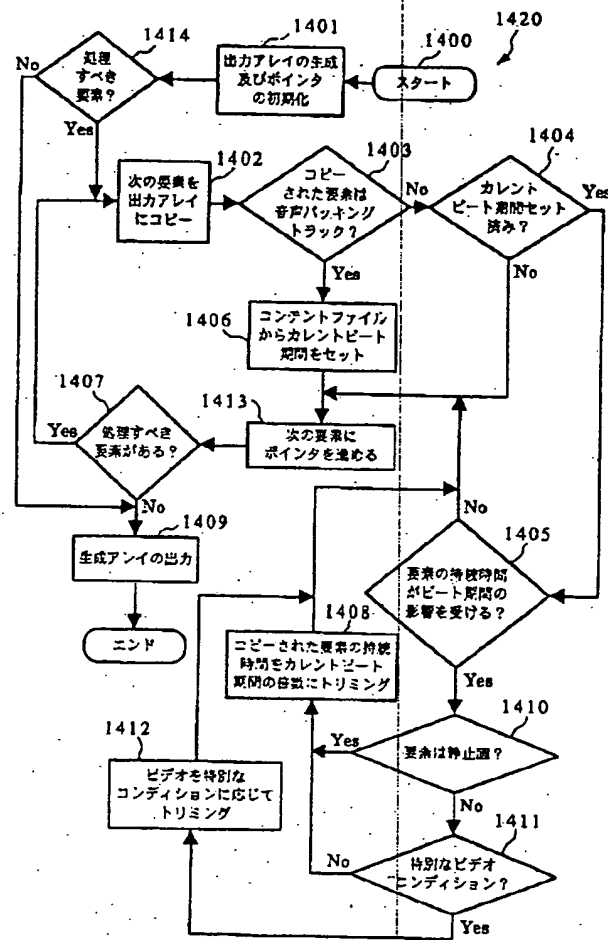
【図2】



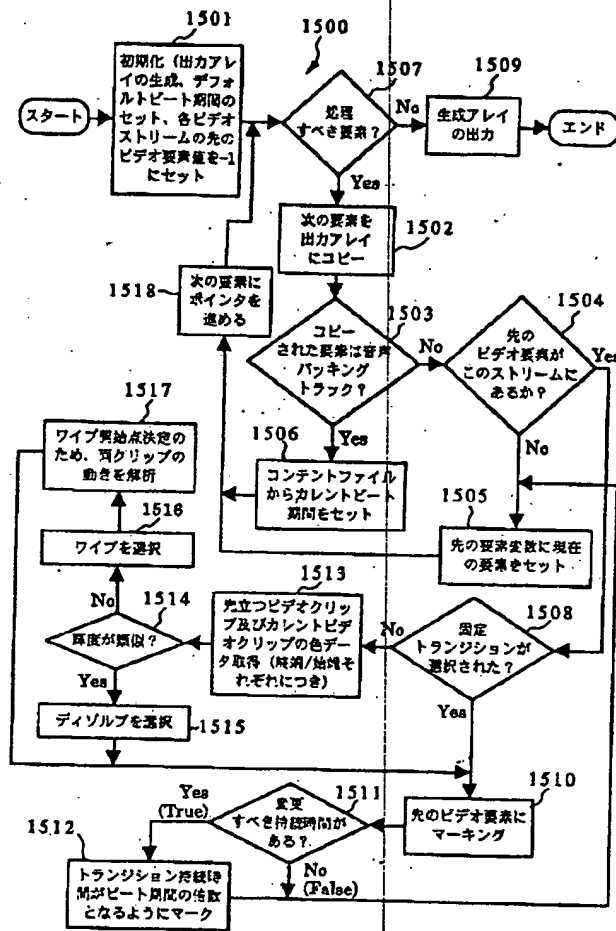
【図12】



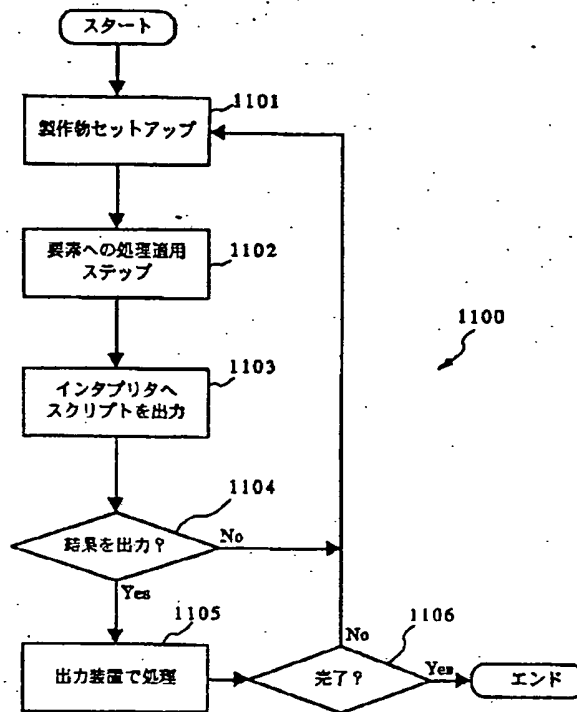
【図3】



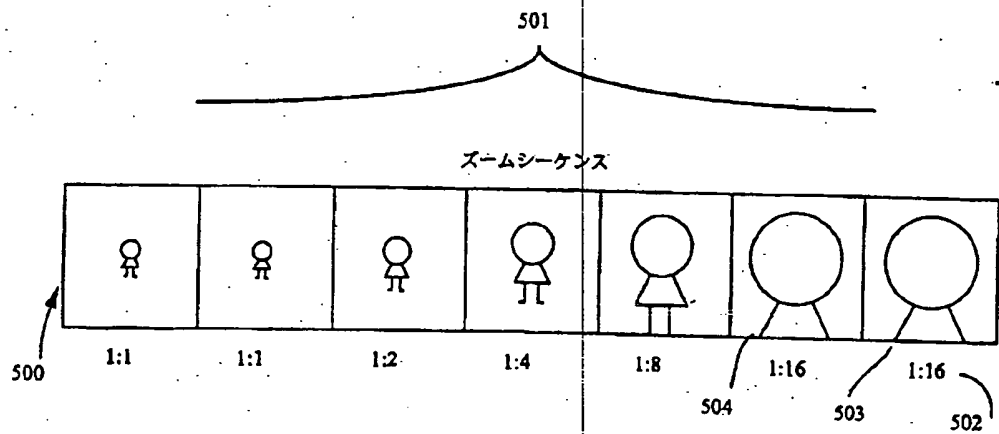
【図4】



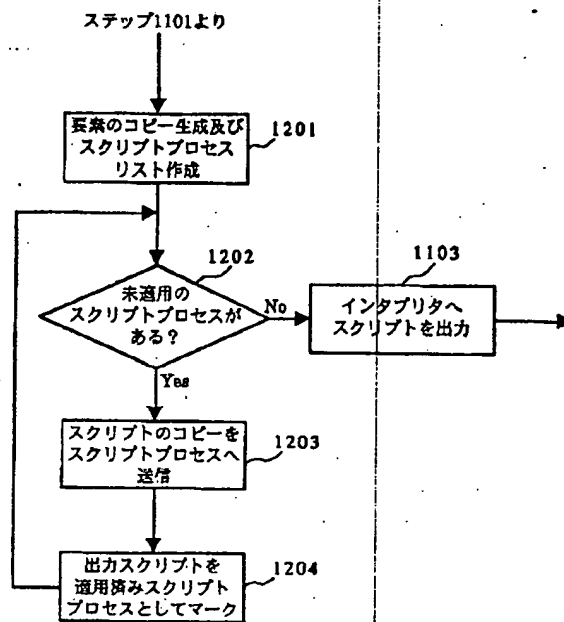
【図5】



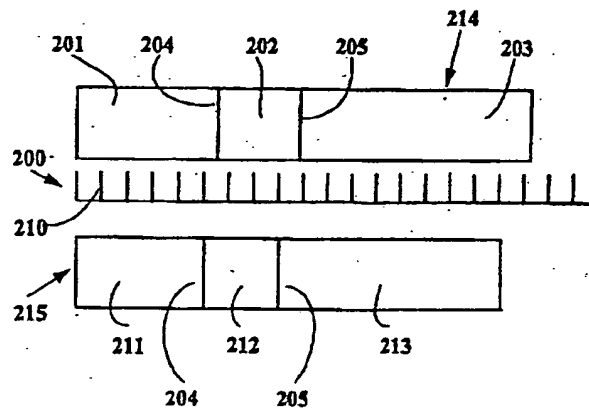
【図9】



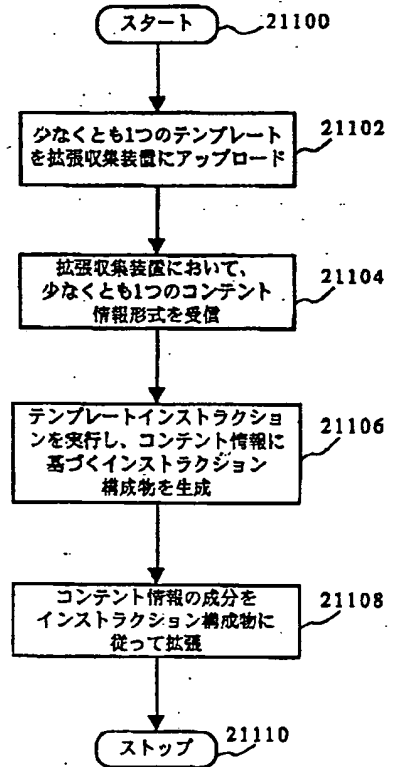
【図6】



【図7】



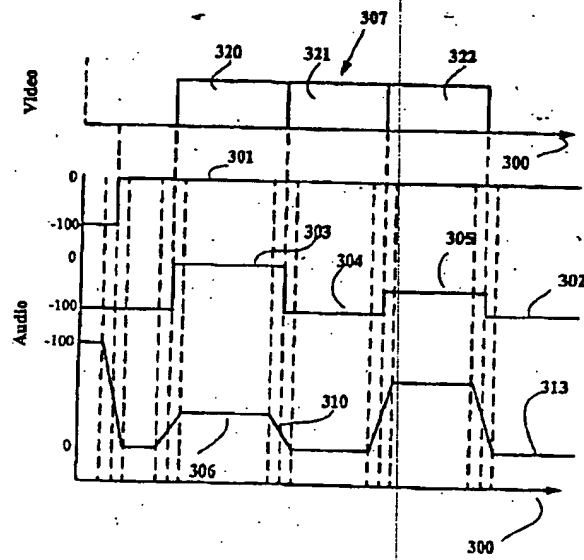
【図28】



(29)

特開平11-234560

【図8】

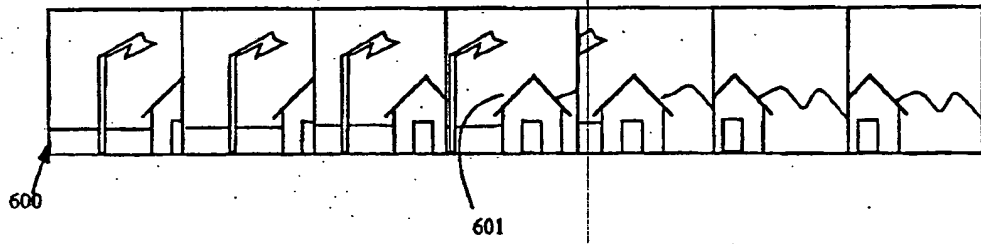


(30)

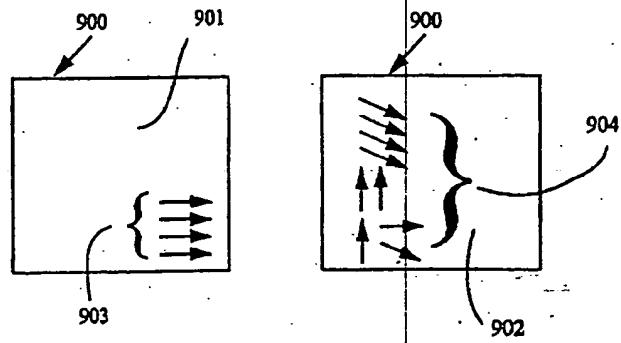
特開平11-234560

【図10】

パンシーケンス



【図13】





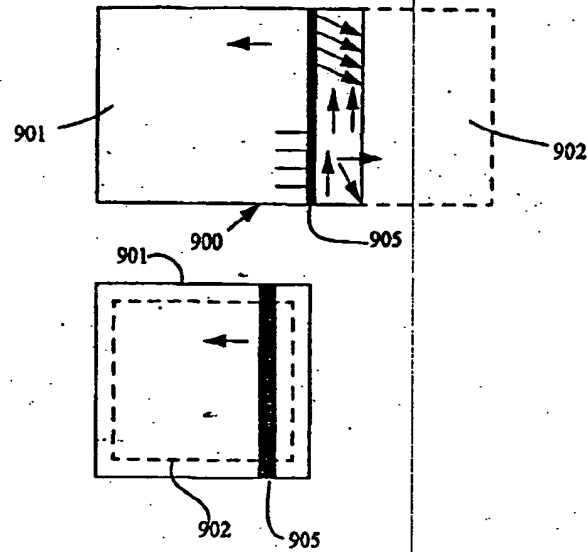
(31)

特開平11-234560

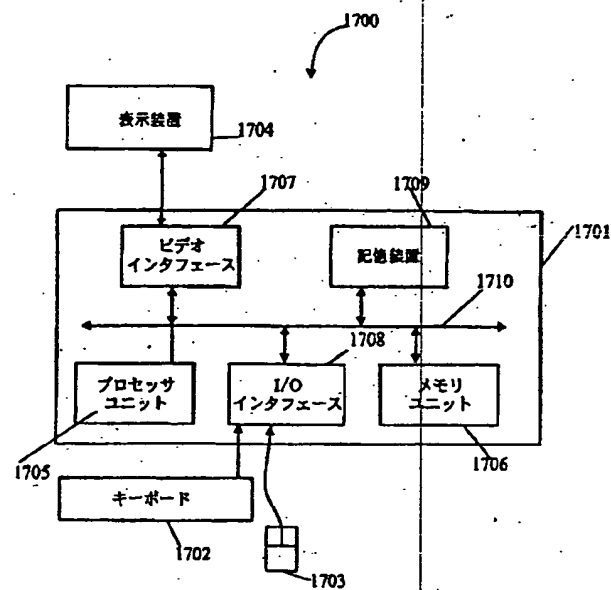
【図11】



【図14】



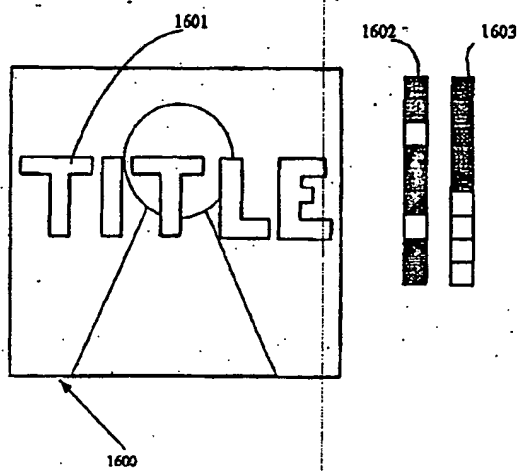
【図17】



(33)

特開平11-234560

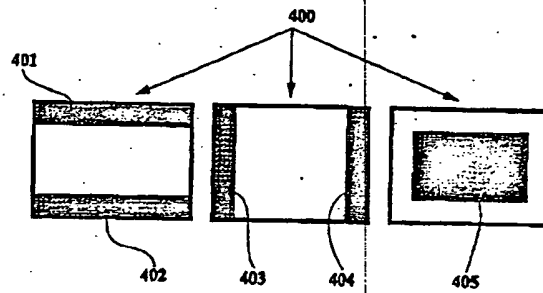
【図15】



(34)

特開平11-234560

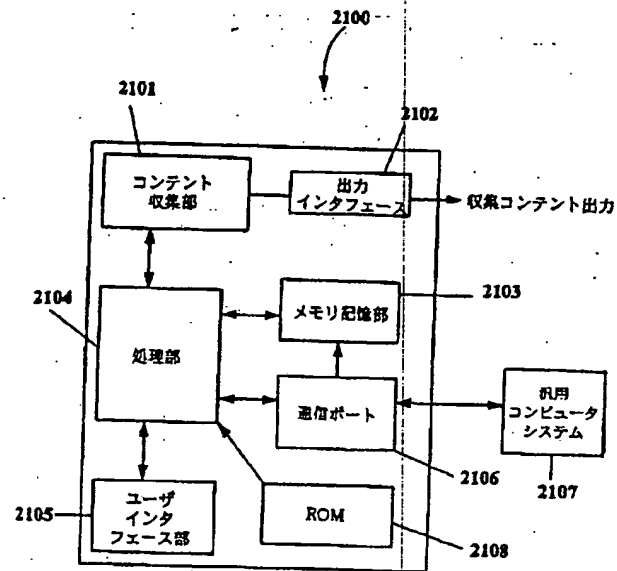
【図16】



(35)

特開平11-234560

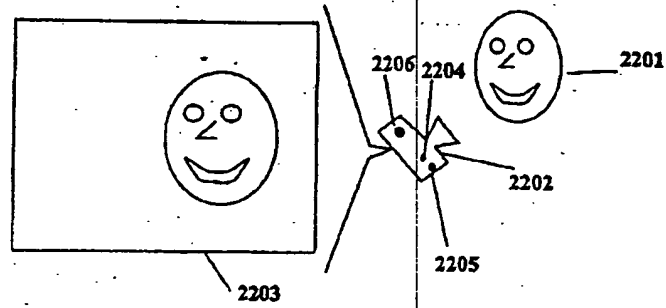
【図18】



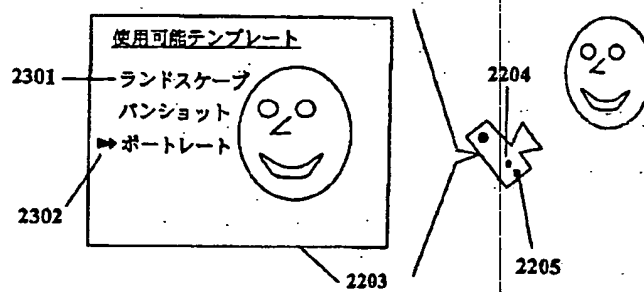
(36)

特開平11-234560

【図19】



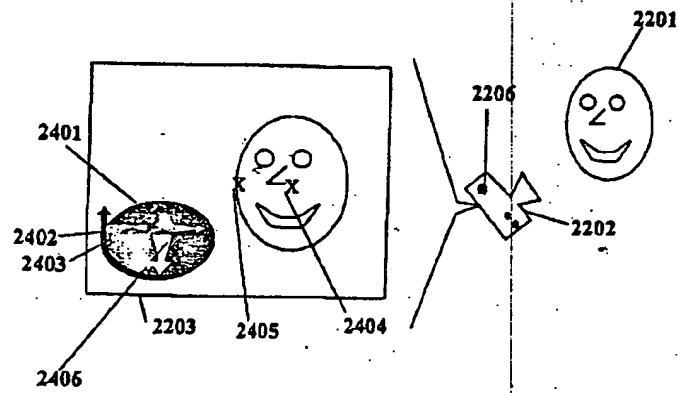
【図20】



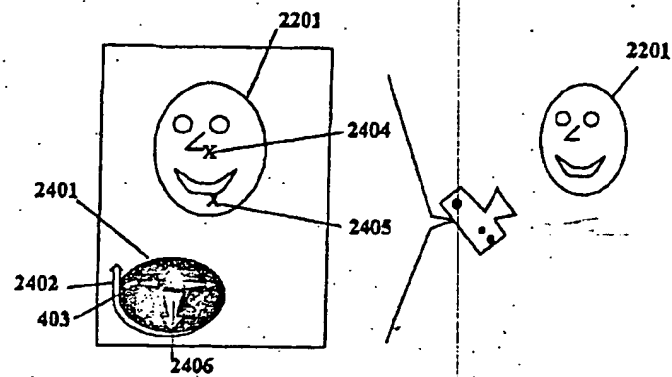
(37)

特開平11-234560

【図21】



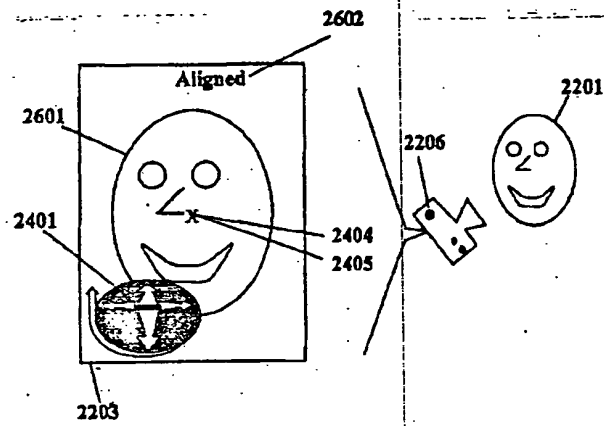
【図22】



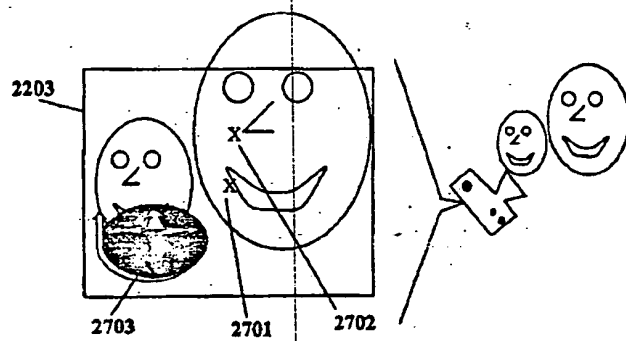
(38)

特開平11-234560

【図23】

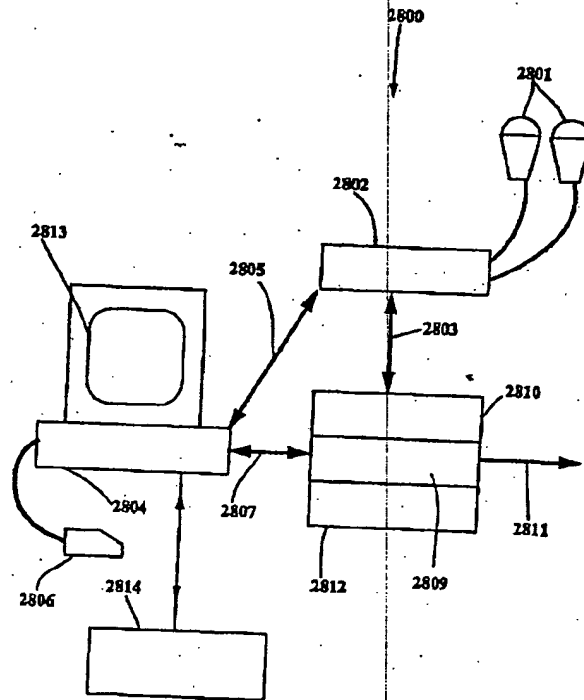


【図24】





【図25】



(40)

特開平11-234560

【図26】

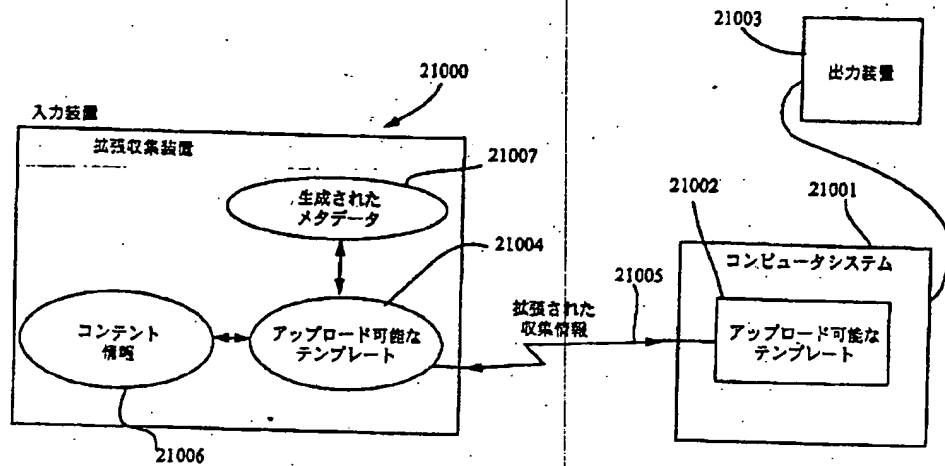
2903 不動産チェックリスト

<input checked="" type="checkbox"/>	前庭
<input checked="" type="checkbox"/>	玄関
<input type="checkbox"/>	台所
<input type="checkbox"/>	風呂
<input type="checkbox"/>	ランドリー
<input type="checkbox"/>	主寝室
<input type="checkbox"/>	客間

2902 { 2901

2900

【図27】



フロントページの続き

(71)出願人 591146745

キャノン インフォメーション システム  
ズ リサーチ オーストラリア プロプラ  
イェタリー リミテッド  
CANON INFORMATION S  
YSTEMS RESEARCH AUS  
TRALIA PTY LTD  
オーストラリア国 2113 ニュー サウス  
ウェールズ州、ノース ライド、ト  
ーマス ホルト ドライブ 1

(72)発明者 ジョン リチャード ウインドル

オーストラリア国 2113 ニュー サウス  
ウェールズ州、ノース ライド、トーマ  
ス ホルト ドライブ 1 キャノン イン  
フォメーション システムズ リサーチ  
オーストラリア プロプライエタリー  
リミテッド内

## 【外国語明細書】

## 1 Title of Invention

## MULTI-MEDIA EDITING METHOD AND APPARATUS

## 2 Claims

1. A computer implemented method for creating a dynamic production for sensory perception by a user, said production including at least one of an audio element and a visual element able to be perceived by the user, said method including the steps of:

(a) extracting from at least one portion of at least one said element content information relating to a perceivable content of said portion; and

(b) creating said production by applying at least one of a predetermined (first) set of production rules to said portion according to the extracted content information.

2. A method according to claim 1 including the further steps, before step (a), of:

(aa) applying to said at least one portion at least one of a predetermined (second) set of production rules to form said content information corresponding thereto; and

(ab) associating said content information with the corresponding said portion.

3. A method according to claim 1 wherein step (b) includes editing said portion to provide a correspondingly altered portion based upon said first set of production rules as selectively applied to said extracted content information.

4. A method according to claim 3 wherein said step (b) further includes examining said extracted content information and said first set of production rules to selectively apply at least one of said first production rules to said portion to form said altered portion.

5. A method according to claim 2 wherein each said element has corresponding rules within each of said first and second sets of production rules and step (aa) includes applying the corresponding production rules to the corresponding said element within said portion.

6. A method according to claim 1 wherein said elements include at least one of audio elements, motion visual elements, and still visual elements and said portions include at least one of a characteristic of said element and a division of said element.

7. A method according to claim 6 wherein said element is an audio element and said portion is at least one of a loudness, a pitch, a tempo or beat, and a delay applied to, said audio element.

8. A method according to claim 6 wherein said element is a visual element and said portion is at least one of a brightness, a colour, a motion, a frame, a segment of a frame, of said visual element.

9. A computerised method of forming a dynamic media production from source media having elements including at least one of audio media and visual media, said method including the steps of:

- (a) receiving said source media;
- (b) analysing at least one said element for at least a portion of said source media with a predetermined set of content rules to extract therefrom content information associated with a corresponding portion of said source media;
- (c) examining the content information of at least one said portion to determine a limited set of production rules applicable to the corresponding said element;
- (d) selectively applying at least one of said production rules to a corresponding one of said elements so as to alter said corresponding one element; and
- (e) combining at least said altered element with other of said media to form said media production.

10. A method according to claim 9 wherein steps (b), (c) and (e) are substantially automatically performed and step (d) includes a user making a selection of said one rule and the corresponding said one element to which the selected rule is to be applied.

11. A method according to claim 10 wherein step (c) includes presenting a representation of said at least one element and the corresponding said limited set of rules to the user in anticipation of the selection of step (d).

12. A method according to claim 9 wherein said source media includes audio and visual elements and said rules include a duration of each of said elements and said method includes the step of modifying a duration of at least one of said audio and visual elements to match a desired duration of the other of said audio and visual elements.

13. A method according to claim 9 wherein said source media includes an audio element and said associated content information includes a beat of said audio element and said method includes altering a duration of said audio element according to said beat.

14. A method according to claim 13 wherein said altered duration is a multiple of a beat period of said audio element.

15. A method according to claim 12 wherein said source media includes a visual element desired for reproduction coincident with said audio altered audio element and said method includes trimming a duration of the visual element in relation to that of said altered audio element.

16. A method according to claim 15 wherein said visual element is trimmed to coincide with the duration of said altered audio element.

17. A method according to claim 16 wherein said visual element comprises a series of frames and said trimming comprises omitting selected ones of said frames.

18. A method according to claim 17 wherein said series of frames comprise individual scenes to be reproduced and said trimming comprises altering a duration of at least one of said scenes so that transitions between said scenes coincide with a multiple of said beat period.

19. A method according to claim 16 comprising the step of examining said series of frames according to a set of omission rules to determine those ones of said frames most appropriate for omission, and then omitting selected ones of the determined frames.

20. A method according to claim 19 wherein said set of omission rule are established so as to identify those ones of said frames that, if omitted, would least impact upon a visual quality of said media production.

21. A method according to claim 20 wherein said omission rules include:

(i) detection those frames least affected by motion relative to others of said frames;

(ii) detecting those frames proximal to a pan or zoom for omitting other ones of said frames; and

(iii) detecting those frames of diminished focus compared to other of said frames and omitting the diminished focus frames.

22. A method according to claim 17 further comprising altering the beat period for reproduction of said audio element to adapt the audio element to said visual element.

23. A method according to claim 17 wherein a duration of said audio element is modified according to a transition between scenes of visual elements, said transitions including a butt edit, a wipe and a fade.

24. A method according to claim 12 further comprising modifying an audio output level associated with said audio element.

24. A method according to claim 9 wherein said content information of said audio element is used to select an alternative audio element having desired characteristics for substitution in said media production.

25. A method according to claim 24 wherein said characteristics are selected from the group consisting of average volume, beat period, and dominant frequencies.

26. A method according to claim 9 wherein said content information of said visual element includes metadata encoded with said visual element with a creation of said visual element.
27. A method according to claim 26 wherein said content information include zoom data regarding a visual image present in said visual element.
28. A method according to claim 9 further comprising the step of printing an image representative of said visual element based upon characteristics of said visual element includes in said content information, said printed image forming an identifier to said user as an significant portion of said visual element..
29. A method according to claim 9 wherein said visual element includes content information related to colour and/or luminance of said visual element.
30. A method according to claim 29 further comprising the step of creating a user generated visual element from an image visual element dependent upon the colour content information thereof.
31. A method according to claim 30 wherein said user generated visual element includes at least one of textual elements and graphic object elements.
32. A method according to claim 9 wherein said content information is generated with a creation of said source media.
33. A method according to claim 9 wherein step (b) is performed substantially simultaneously with step (a).
34. A method according to claim 29 wherein said content information is used to determine one of a wipe or a dissolve between adjacent portions of said visual element.
35. Apparatus configured to perform the method of any one of claims 1 to 34.



36. A computer program product including a computer readable medium incorporating a series of computer implementable steps arranged for performing the method of any one of claims 1 to 34.

37. A media production system comprising:  
a media input arrangement for receiving source media including at least one of audio media and visual media

analysing means for analysing said source media to extract therefrom content information associated with a corresponding element of said source media;

examining means for examining the content information of at least one said element to determine a limited set of production rules applicable to the corresponding said element;

command means for presenting to a user of said system at least said set of production rules and corresponding representation of said element and for receiving user input commands of selected ones of said rules to be applied to corresponding said elements;

modifying means for applying said selected rules to a corresponding one of said elements so as to alter said corresponding one element; and

combining means for combining at least said altered element with other of said media to form said media production.

38. A system according to claim 37 wherein said source media includes audio and visual elements and said rules include a duration of each of said elements for which a duration of at least one of said audio and visual elements modifiable to match a desired duration of the other of said audio and visual elements.

39. A system according to claim 37 wherein said source media includes an audio element and said associated content information includes a beat of said audio element and said modifying means acts to alter a duration of said audio element according to said beat.

40. A system according to claim 39 wherein said altered duration is a multiple of a beat period of said audio element.

41. A system according to claim 39 wherein said source media includes a visual element desired for reproduction coincident with said audio altered audio element and modifying means acts to trim a duration of the visual element in relation to that of said altered audio element.

42. A system according to claim 41 wherein said visual element is trimmed to coincide with the duration of said altered audio element.

43. A system according to claim 42 wherein said visual element comprises a series of frames and said trimming comprises omitting selected ones of said frames.

44. A system according to claim 43 wherein said series of frames comprise individual scenes to be reproduced and said trimming comprises altering a duration of at least one of said scenes so that transitions between said scenes coincide with a multiple of said beat period.

45. A system according to claim 43 wherein said examining means examines said series of frames according to a set of omission rules to determine those ones of said frames most appropriate for omission, and said modifying means is operable to omit ones of the determined frames as selected by said user.

46. A system according to claim 45 wherein said set of omission rule are established so as to identify those ones of said frames that, if omitted, would least impact upon a visual quality of said media production.

47. A system according to claim 45 wherein said omission rules include:

- (i) detection those frames least affected by motion relative to others of said frames;
- (ii) detecting those frames proximal to a pan or zoom for omitting other ones of said frames; and
- (iii) detecting those frames of diminished focus compared to other of said frames and omitting the diminished focus frames.

48. A system according to claim 40 further wherein said modifying means is operable to alter the beat period for reproduction of said audio element to adapt the audio element to said visual element.

49. A system according to claim 39 wherein said modifying means is operable to alter a duration of said audio element according to a transition between scenes of visual elements, said transitions including a butt edit, a wipe and a fade.

50. A system according to claim 39 wherein said modifying means is operable to alter an audio output level associated with said audio element.

51. A system according to claim 37 wherein said content information of said audio element is used to select at least one alternative audio element having desired characteristics for substitution in said media production.

54. A system according to claim 51 wherein said characteristics are selected from the group consisting of average volume, beat period, and dominant frequencies.

55. A system according to claim 37 wherein said content information of said visual element includes metadata encoded with said visual element with a creation of said visual element.

56. A system according to claim 55 wherein said content information include zoom data regarding a visual image present in said visual element.

57. A system according to claim 37 further comprising means for printing an image representative of said visual element based upon characteristics of said visual element includes in said content information, said printed image forming an identifier to said user as a significant portion of said visual element.

58. A system according to claim 37 wherein said visual element includes content information related to colour of said visual element.

59. A system according to claim 58 further comprising means for creating a user generated visual element from an image visual element dependent upon the colour content information thereof.

60. A computer readable medium incorporating a series of computer implementable steps arranged for forming a dynamic media production from source media having elements including at least one of audio media and visual media; said computer program product including:

a first module for receiving said source media;

a second module for analysing at least one said element for at least a portion of said source media with a predetermined set of content rules to extract therefrom content information associated with a corresponding element of said portion of source media;

a third module for examining the content information of at least one said portion to determine a limited set of production rules applicable to the corresponding said element;

a fourth module for selectively applying at least one of said production rules to a corresponding one of said elements so as to alter said corresponding one element; and

a fifth module for combining at least said altered element with other of said media to form said media production.

### 3 Detailed Description Of Invention

#### Field of the Invention

The present invention generally relates multi-media editing and, in particular, a method of creating multi-media productions that uses content information of elements of the production to guide in a construction of the multi-media production.

#### Background Art

Combining still images, text, videos, sound and other multimedia elements into a production is a complex task that requires extensive knowledge in many diverse areas. Video editing is a well established field in which numerous types of elements are combined to form a final production. The types of elements often used in the construction of a video production are visual, audible and textual. Visual elements include such things as video footage, still shots and animation. Audible elements include such things as backing tracks, sound

effects, commentaries and recorded audio of the video clips. Textual elements include such things as titles, credits and subtitles.

These elements have various properties that can be altered and effects that can be applied. The elements can then be combined in many different ways to form a final product and for each combination there is a large amount of knowledge that is relevant to the choices that must be made.

For many years, this extent of production was available only to TV stations and large production houses. With the advent of home video cameras, personal computers, inexpensive video capture cards and video editing software, it is now possible for many more people to create their own video productions.

However, the knowledge needed to make good choices must be learnt by the person creating the production. For many people who only wish to put together a few simple productions the learning curve is quite steep and poor quality productions are created due to a lack of informed decisions. Similar difficulties arises in the case where it is desired to create a production made up of digitised stills and music, stills and text, video material alone, or sound material alone.

Thus there exists a need for an apparatus, system and/or method for a guided creation of a production which provides assistance and guidance to a user creating such a production.

#### Summary of the Invention

In accordance with one general aspect of the present invention there is provided a computer implemented method for creating a dynamic production for sensory perception by a user, said production including at least one of an audio element and a visual element able to be perceived by the user, said method including the steps of:

- (a) extracting from at least one portion of at least one said element content information relating to a perceivable content of said portion; and
- (b) creating said production by applying at least one of a predetermined (first) set of production rules to said portion according to the extracted content information.

Preferably, the method includes the further steps, before step (a), of:

- (aa) applying to said at least one portion at least one of a predetermined (second) set of production rules to form said content information corresponding thereto; and

(ab) associating said content information with the corresponding said portion.

In accordance with another general aspect of the present invention there is provided a computerised method of forming a dynamic media production from source media having elements including at least one of audio media and visual media, said method including the steps of:

- (a) receiving said source media;
- (b) analysing at least one said element for at least a portion of said source media with a predetermined set of content rules to extract therefrom content information associated with a corresponding portion of said source media;
- (c) examining the content information of at least one said portion to determine a limited set of production rules applicable to the corresponding said element;
- (d) selectively applying at least one of said production rules to a corresponding one of said elements so as to alter said corresponding one element; and
- (e) combining at least said altered element with other of said media to form said media production.

Other aspects of the invention including apparatus, systems and computer program products will become apparent from an understanding of the following description and of the claims appended hereto.

#### Detailed Description

Fig. 1 shows a block diagram of an overview a preferred embodiment of a Multimedia Editing device 100. The Multimedia Editing device 100 includes a processing unit 101, a storage unit 102, a user interface unit 103, and communications port and corresponding bus 104. The storage unit 102 stores application software code used by the processor unit 101 to control and respond to the user interface unit 103. The storage unit 102 is also used to store various rules used in the creation of a production and one or more elements that are to be used in the production. The term production is used here in the generic sense, because a production could be a video production, a musical production, a series of still photographs presented as a slide show or a combination of these. Elements, as described previously, include visual elements, audio elements and/or textual elements which can be described, quite generally, as components of a production. The visual elements may include video footage, video clips, that is,

a portion of a video footage; a single video frame, an entire movie, computer clip-art, photographs, and created images such as those used in video animation. The audible elements may include sound recordings, sound effects, music and musical scores, single and multiple tones or notes. Textual elements can include titles and credits of a production, advertising such as logos or slogans, and/or other indicia.

In the overview of the preferred embodiment the communication port and corresponding bus 104 is used to transfer data to and from the storage unit 102. The data is typically a digital representation of the elements previously described. The communication port and corresponding bus 104 is use for other purposes as well, for instance downloading or uploading the application software code. An example of a bus 104, and corresponding interface, is that defined by the IEEE 1394 high-speed serial bus standard.

The processing unit 101, aided by the application software, performs and analysis of the elements to extract content information. The extracted context information is stored in the storage unit 102 and is associated with the corresponding element from which the content information was extracted. The content information is specific to the various different types of elements, (eg., visual, audio, textual) and as such the analysis performed by the processing unit 101, in conjunction with the software, will vary accordingly. The analysis performed on the various elements to extract, or in some cases generated, the content information will be described hereinafter under a corresponding subsection.

The analysis for extracting the content information can be done as an off-line process, that is, at some time after the various elements have been downloaded to the storage unit 102. However, it is envisaged that at least some, if not all, of the analysis could be done on the fly as the data of the various elements is downloaded 104 to the storage unit 102. In the preferred implementation, the analysis of the various elements may be performed at least in part by a recording or capture device which is used to obtain the element. For example, a digital video camera (DVC) can be modified to store or record not only image data and sound data, but also metadata such as date, time, number of video frames take etc. An example of a DVC for capturing selected elements which can be used by the preferred embodiment is described below as an "Enhanced Gathering Apparatus".

Referring to Fig. 2, there is shown a block diagram representing an input script 1300 comprising a plurality of visual and audio elements for generating a production. The input script 1300 of Fig. 2 is an example of a production which is a video presentation, however input scripts for other forms of productions such as audio-productions, still frame \_slide show type\_ presentation, productions such as printed matter, (ie. a book, catalogue etc.), are also possible. The visual elements of the script 1300 include video segments 1302, 1304 and 1305, preferably stored in digital form, and a digital still image 1303 which can be a video segment having substantially one image reproduced a number of times so as to provide the appearance, when rendered to a display, of a still image or a single frame image paused for a pre-determined duration. The audio elements comprise audio backing tracks 1301 and 1306. Typically the audio backing tracks include musical tracks and/or sound effects, and in some cases are audio tracks in addition to an audio track associated with a video segment. Throughout this description a reference to a video segment, unless specified to the contrary, is a reference to a video segment including an associated audio track, where the audio track is available. For example, video footage usually comes with a sound recording associated with the video footage, like a person's voice or background noise. This audio information is assumed to be part of a video segment for the purpose of the description of the preferred embodiment, although it will be appreciated that embodiments may be practiced using only one of visual or audio source information.

Input scripts may be rendered (reproduced) either sequentially or in parallel, as the situation dictates or as need arise. The input script 1300 is preferably rendered sequentially. However, as in most multimedia implementations, audio elements are typically rendered in parallel via different channels to the visual elements. In the arrangement of the input script 1300, a first audio backing track 1301 is commenced and outputted to an audio device concurrently with a sequential output of visual elements 1302, 1303 and 1304 to a display device. At the end of the visual element 1304, a transition occurs from audio backing track 1301 to the next sequential audio backing track 1306, and a further sequential visual element 1305 is displayed concurrently with the audio backing track 1306. A synchronicity between a commencement of the visual element and an audio backing track is desired, and preferably associated timing information is encoded with each element, whether the element is visual, audio or textual. Extraction of content information may be performed "live", for example



whilst recording the visual image and/or the sound source. This may be achieved substantially automatically or semi-automatically with user input. In some instances, content information may be generated after recording takes place under manual control or through automated post-processing. Preferably each element further includes:

- \* information that directs a processing application to a source data and content data associated with the element; and
- \* supplementary information such as duration of an element, and an offset start time relative to source data.

Referring to Fig. 3, there is shown a flow chart of a beat synchronisation process 1420 in accordance with the preferred embodiment. The beat synchronisation process 1420 receive input in the form of a plurality of elements (ie. an input array of elements) in a pre-determined sequence or order. An output of the beat synchronisation process 1420 is an input script substantially of the form of the input script 1300 of Fig. 2. However each element of the input array of elements may or may not be altered by the beat synchronisation process 1420.

Initially at step 1400 the beat synchronisation process 1420 receives as input an array of elements (visual, audio etc.) of a production, preferably in a predetermined order. Next an empty output array is created at step 1401 having a substantially similar number of available entry slots as the number of elements of the input array. In this creation step 1401, a pointer is also initialised to a starting point of a first element of the input array. Control is then passed to a next step, a decision step 1414, which decides if there are elements in the input array to be processed. If the result of the decision step 1414 is false, then a current output array is outputted 1409 in a format represented by the input script 1300. Otherwise, if there exists at least one element, the decision step 1414 returns true and proceeds to a copy step 1402 where a current element of the input array (source array) is copied to a slot of the output array. The copy step 1402 is followed by a next step in the flowchart 1420, an audio element decision step 1403, which returns a true value if the current element is an audio backing track, otherwise the step 1403 returns a false. At a true result from the audio element decision step 1403 the process enters a read input step 1406 which reads from a supplementary information portion of the current element the beat period, and sets a current beat period accordingly. This step may include calculating the beat period. The pointer is advanced at step 1413 to a next element of the input array. At an otherwise false result from the audio element

decision step 1403, an additional decision step 1404 is entered which determines whether or not a current beat frequency has been set. The additional decision step 1404 return a true value when a current beat period has been set, and false otherwise. A false decision at step 1404 results in the pointer advancing at step 1413 to point to the next element of the input array. A true result from the additional decision step 1404 has the process 1420 enter yet a further decision step 1405 which determines whether or not the current element is affected by the current beat period. In the event that the current element is not affected by the current beat period, the pointer is advanced at step 1413 to the next element. A true result, from the decision step 1405 for determining whether or not the current element is affected by the beat period, leads to a further a decision step 1410 for deciding whether or not the current element is a still shot sequence. At this decision step 1410, if the current element is a still shot then a trimming step 1408 is performed on the current element, the current element being the current element of the output array, that is, the element copied at process step 1402. Following the trimming step 1408 the pointer is advanced at step 1413 to the next element.

A false return from decision step 1410 (ie. the current element is not a still shot), leads to a decision step 1411 which determines if the current element, not being a backing track or still image, includes predetermined video conditions such as a zoom shot sequence, a pan shot sequence or other predetermined video special effects. No video conditions being identified (determined) for the current element, a trimming step 1408 is performed on the current element. On each occasion that the trimming step 1408 is performed on an element (visual element). The element is trimmed to substantially a multiple of the current beat period. As previously described, following the trimming step 1408 the pointer is advanced 1413 to the next element (ie. a next input element, to be copied to the output array, if any). If a video condition is identified, at step 1411, a true is returned from the decision step 1411 and the current element is trimmed at step 1412 in accordance with predetermined rules defined for each video condition. Examples of the predetermined rules applied are described hereinafter. The process flow following the trimming step 1412 leads to the pointer advancing step 1413 previously described. Once the pointer is advanced to the next element of the input array, a decision step 1407 checks whether or not there are further elements in the input array. If so, the process is looped back to the copy step 1402 and the steps previously described subsequent to the copy step 1402 are processed until the decision step 1407, checking for further input elements returns a false, in

which case the output array is outputted 1409 ending the beat synchronisation process 1420 until it is re-evoked. Although the above describes trimming the video, similar processing may be used to either stretch or compress the video so as to match the beat frequency.

Referring now to Fig. 4, there is shown a flow chart of a transition generating process 1500 in accordance with the preferred embodiment. The transition generating process 1500 determines a type of transition to be used between video elements and determines the duration of a transition. The transition is not rendered until a final production is rendered to an output display device.

Initially, the transition generating process 1500 receives as input an array of elements. Each element in the input array includes information of a corresponding channel, or channels, to which it will be rendered. At a first step 1501 in the process 1500, an empty output array is created with substantially the same number of slots as the number of elements of the input array. At this step 1501 initialisation procedures are also performed. For example, a source pointer is set to point to a first element of the input array, a destination pointer is set to point to a first available slot of the output array, and a beat period variable is set to a default value until a new value for the beat period is determined. Further, for each video channel, a variable (ie. "PrevEle") is supported which points to a previous video element. Initially, since no previous video elements are available, the variable PrevEle is assigned a value indicating no previous video elements. Following initialisation, control in the process 1500 is transferred to a decision step 1507 which determines whether or not there are elements in the input array to be processed. If no elements are to be processed, the decision step 1507 returns a false and current values of the output array are outputted in a form of an input script substantially as illustrated by Fig. 2.

If the input array is an empty array then an empty input script will result (ie. no input script). However, if there are elements of the input array to be processed, decision step 1507 will return true which leads the process to a copy step 1502. At the copy step 1502 a current input (array) element is copied to an available slot of the output array. Subsequent to the copy step 1502 a decision step 1503 is entered which determines whether or not a current element (copied element) is an audio backing track. If the current element is an audio backing track, the current beat period is set to the extracted value in step 1506 and incremented in step 1518 to a next element of the input array. Otherwise, at

decision step 1503, the current element is not a audio backing track and a next step 1504 is entered which determines whether or not a previous video element on the current element's video channel, for which a transition is required to the current video element, exists. If no previous video element exists on the current video channel, a set PrevEle step 1505 is entered, which sets the PrevEle variable to point to the current element and then passes to the increment step 1518. If a previous video element (ie. to the current (video) element), is available, the process flow enters a decision step 1508 which checks whether or not a fixed transition style (eg. a butt edit transition) has been selected by a user of the Multimedia Editing device 100. Where a fixed transition has been selected, a marking step 1510 is entered which marks the previous video element for transition to the current video element.

If no fixed transition is selected, the decision step 1508 passes to an extract colour data step 1513. Preferably luminance data is extracted from a tail end of a previous video element and a front end of the current video element. A next step 1514 is entered which decides on the basis of the extracted colour data whether to choose a dissolve transition, selected in step 1515, or to choose a wipe transition, selected in step 1516. In the case of a wipe transition, an additional step 1517 is used to analyse motion vectors of both a previous and the current video element to determine details of the wipe transitions. Such wipe transition details include starting point, speed of wipe etc.

Following on from the additional step 1517 of the chosen wipe transition or from step 1515 for choosing a dissolve transition, both return to the marking step 1510 previously described. At the marking step 1510 the previous video element is marked for transition to the current video element in accordance with the chosen transition (ie. Dissolve, wipe, fixed etc.). Following the marking step 1510, a decision step 1511 is entered which determines if the chosen transition has a duration length longer than a predetermined duration length. Optionally, the decision step 1511 checks for fixed transitions either in addition to or instead of the predetermined duration length. At decision step 1511, a true value indicates a duration longer than the predetermined duration length, in which case step 1512 marks a duration length of the chosen transition to be a multiple of the current beat period. Otherwise, decision step 1511 returns a false indicating a shorter duration length than the predetermined duration length. Such a situation leads the flow chart back to the set PrevEle step 1505.

Further, once the determining step 1512 has marked the duration length of a chosen transition, the flow is also returned to the set PrevEle step 1505.

Again, as previously described, the set PrevEle step 1505 sets the PrevEle variable to point to the current video element and returns to the increment to the next input element step 1518. This step 1518 increments the source pointer and continues for a next input element as hereinbefore described.

The transition generating process 1500 terminates when no further input elements are available for processing, in which case the current elements of the output array are outputted substantially in the form of the input script 1300 of Fig. 2.

Fig. 5 shows an overview 1100 of the editing process in accordance with the preferred embodiment of the present invention. Initially, a production set-up step 1101 occurs allowing the user to select desired visual, audio and/or textual elements for a production. Next, scripting processes 1102 are applied to the user selected elements. These scripting processes 1102 include the beat synchronisation process and the transition generating process previously described with reference to Figs. 3 and 4 respectively. Without departing from the spirit and scope of the present invention, other scripting process can be applied. For example, an editing process that edits by manipulating picture elements of a frame of a video clip or still image. The output of a scripting process is preferably an input script substantially as described with reference to Fig. 2.

Input scripts, resulting from the scripting processes 1102, are processed at a next step 1103 where each input script is interpreted by a script interpreter. A decision step 1104 is entered which decides for each input script whether or not an output from an input script is to be rendered to an output device. If the answer to the decision step 1104 is no, the flow process, described with reference to Fig. 5, is looped back to the production setup step 1101 for additional desired selections by the user. However, if the answer is yes at step 1104, that is an output is to be rendered to an output device then each input script, comprising elements of a production, is then rendered at step 1105 to a corresponding output device. For example, for an input script comprising audio elements and video elements, the audio elements are rendered to an audio output device and a video portion of the video elements to a display device, while a sound track associated with the video element is also rendered to the audio output device. Although Fig. 5 refers to a single interpreter, typically one interpreter

will be provided for each type of source data and for the corresponding output device (eg. audio, visual).

A final decision step 1106 checks if the production is finished, or loops back to the production setup step 1106 if it is to continue allowing user selections. the loop back allows for a preview of the production by the user. At the production setup step 1106, the user is also permitted to select which scripting process is, or is not, to be applied. For example, the user may desire that no beat synchronisation be performed and so this option is switched off. Further, the user can also select desired output devices.

Turning to Fig. 6, there is shown the apply scripting processes step 1102 of Fig. 5 in more detail. At a first step 1201 a copy of user selected elements is made, to avoid altering of original elements, and a list of scripting processes to be applied is created. Next, a decision step 1202 determines whether or not all scripting processes on the list have been applied. That is, the decision step 1202 determines whether the list of scripting processes contains any scripting processes not yet applied. If the answer is no that is, all scripting processes have been applied, then the input scripts generated by the scripting processes are sent to the script interpreter as described with reference to Fig. 5. However, if not all of the scripting processes (on the created list) have been applied, at decision step 1202, then the copied elements are passed to a corresponding scripting process 1203. For example, if a beat synchronisation process is to be applied to a current copied elements then the elements are directed to the beat synchronisation process. A next step 1204 is to mark off the list an applied scripting process and loop back to the decision step 1202 to assess whether another scripting process is to be applied. When no more scripting processes are to be applied a final generated input script is sent to the script interpreter for processing 1103 (see Fig. 5).

A predetermined set of rules may now be defined to guide the creation of a production from a collection of elements, however this set of rules should not to be seen as limiting and is only offered by way of example. The examples of the rules that can be used in conjunction with the preferred embodiment, are described below, however the type of rule depends on the final production effect to be achieve. It will therefore be appreciated by those skilled in the art that a different combination and/or variety of rules can be applied without departing from the spirit or scope for the present invention. For example, a rule can be created that designates that titles or text information to be rendered, to a video

screen, as overlay on a video clip, can be overlaid on a portion of the video screen where it is determined that the average motion, over the entire video clip, is least. The type of rule also depends on the type of content information to be extracted from the different and various elements. The proceeding sections describe different content information extracted for the various elements and associated rules which are used to create, or assist in the creation of, a production.

#### Beat Synchronisation

Referring to Fig. 7, there is shown an example of automatic beat synchronisation between an audio tract 200, having a beat 210 represented thereon, and a video clip (or video footage) 214. Typically, the audio track 200 for which a beat is sought and used to synchronise with a video clip 214 is a musical piece. The beat is preferably a musical beat of the musical piece represented by the audio track 200, however any substantially regular frequency extracted from the audio track 200 may suffice as the beat for the purpose of synchronisation.

An example of a technique for extracting the beat to allow synchronisation of video clips will now be described. An input audio signal representing the audio element of a production and in particular a section of the audio signal for which a beat is desired, is analysed through a bandpass filter to restrict frequencies of the audio signal to a predetermined range. The restricted frequency signal is passed through a lead edge detecting circuit that detects a lead edge of the restricted frequency signal. Time intervals between each detected lead edge, for the duration of audio signal, for which a beat is sought, are recorded and then an average time interval is calculated. This average time interval can then be used as the beat to synchronise the musical piece to a desired video clip.

The video clip 214 of Fig. 7 represents an untrimmed video footage before it is trimmed to be synchronised with the beat of the audio track 200. In this example the untrimmed video footage 214 comprises three sections 201, 202 and 203 which represent predetermined scenes of the video footage 214. Transitions between the scenes, in this example, are formed by butt editing the scenes together. That is, three sections, each section typically comprising a plurality of frames, have been joined together end on end.

Transitions between sections 201, 202 and 203 of the untrimmed video footage are represented in the illustrated example of Fig. 7, by vertical lines 204 and 205. The transitions of the untrimmed video footage 204 do not coincide with a beat 210 of the audio track 200. The first section 201 of the untrimmed video footage 214 is trimmed so that the resulting section 211 has a transition 203,

to a next section 212, coincident with a beat 210 of the audio track. Similarly the remaining sections 202 and 203 are trimmed to result in sections 212 and 213 respectively, so each transition between sections 211, 212 and 213 occur in synchronism with the beat 210 of the audio track. The trimmed sections 211, 212 and 213 together form a video clip (footage) where transitions between events or changes in scene are synchronised to the beat 210 of the audio track 200. As described in Fig. 3 at step 1413, when trimming video footage and video frames are removed from each section trimmed, such frames can be removed from any portion of the section. However it is preferred that a content of the section is evaluated for predetermined effects and the frames removed are removed so as to cause least disruption to the eventual viewing of the production. For example, if a section of video footage has a zoom or pan then it may be desirable to remove frames which are not intermediate a zoom or pan. Another example would be to assess from a video footage which portion has least motion occurring, that is, the portion of video footage which is essentially a still shot, and remove a frame substantially from the portion of least motion. Alternately, frames which are out of focus in a predetermined section of video footage may be removed.

Further when synchronising the video footage 214 to the audio track 200, it is permissible to process the audio track 200 to increase or decrease an average time interval between beats to the extent that the audio track is not substantially distorted, so as to fine tune a synchronisation between the audio track 200 and the video footage 214. As mentioned previously it may also be appropriate to process the video to speed-up or slow-down without editing.

#### Transition Speed and Timing Rule

Generally a video frame rate (frames/second) is substantially constant, therefore a predetermined number of frames of a video footage substantially defines a specified interval in time. A time interval and hence a transition speed between a first section and a second section of a video footage will then depend on a number of frames of the video footage used in a transition. A transition can be a simple transition from one scene to another scene of the video footage, resulting from a butt edit of two section as described with reference to Fig. 7. However, preferably a transition is purposely designed to provide a special effect such as a dissolve transition from one scene to a next scene. Other transition types include a wipe, and a fade in which allow a change of scene with a predetermined special effect. A timing and speed of such transitions are synchronised with a predetermined audio track in accordance with the preferred



embodiment and in a substantially similar manner as that described above under beat synchronisation, excepting that video frames can be duplicated and added, as well as removed, to effect a smooth synchronised transition. A timing and transition speed rule is applied to start a selected transition at a first predetermined beat marker and ensures, by allowing a predetermined number of frame to effect the transition, that an end frame of the transition is coincident with a second predetermined beat marker. Audio transitions may be achieved by arranging respective beats to be coincident during the crossover.

#### Audio Level Adjustment

Referring now to Fig. 8 there is illustrated a representation of audio level information 301 of an audio backing track and a sequence of video clips 320-322 which are to be associated with the audio backing track. A horizontal axis 300 of Fig. 8 is a time lapsed axis 300 and a vertical axis represents audio levels measured in decibels (dB) (note: Fig. 8 is not drawn to scale). A familiar feature of video cameras is their capability to record sound as well as video. However, when putting together a production it is usually desired to include a further audio channel which is an audio backing track. Additional backing tracks that provide background audio are common feature in professionally prepared video productions, such as a professionally edited wedding video, however such video editing is a laborious task which is painstakingly put together by a video editor skilled in the art.

In the preferred embodiment of the present invention, transitions between video clips 320 to 322 are automatically detected and the audio levels are also automatically detected. For example an audio level 303 corresponding to video clip 320 is detected to be zero decibels (0dB), an audio level 304 corresponding to video clip 321 is detected to be absent and so is set to -100db and for video clip 322 an audio level 305 is detected at -3db.

When rendering a production video, a final attenuation waveform 313 is generated with predetermined ramping 310 at each point 311 corresponding to a video transition from one clip to the next. The final attenuation waveform 313 is used preferably to control a level setting for one or more musical scores used as backing track to the video footage 307. Further, the attenuation waveform may be applied to the backing track so it remains at a fixed distance from audio associated with the video. In addition, the user sets a desired audio difference level between the video camera recorded sound and the audio level of the backing

track so that the recorded sound is not obscure by the backing track if this is so desired.

#### Music Matching

In the preferred embodiment, a first music track is selected and can be used to find and select from a plurality of music tracks, predetermined music tracks which are substantially similar to the first music track. These predetermined musical tracks can, optionally, be used together with or instead of the first music track selected. A technique determining substantially similar musical tracks, from a plurality of music tracks, is based on characteristics of audio signal of each music track in comparison with characteristic of the first music track selected. The characteristics include such features as average volume, beat, and dominant frequencies for example, either singularly or in combination. An example of how this music matching rule is applied in connection with a video production is now described. A template for a video production is supplied, say by a manufacturer, with a first music track selected. A user can offer a plurality of alternative music tracks which the user prefers over the first music track supplied with the template. The music selection rule selects based on one or more predetermined characteristics of the first track, a most appropriate music track from the plurality of alternate music tracks.

#### Visual Content and Motion

In video footage, a plurality of significant visual events may be relied upon to identify segments of the video footage for which predetermined special effects are, optionally, applied. For example, motion detection techniques, known in the art, are applied to the video footage to guide the type of transition that is preferred between a video segment (clip) and a next video segment.

Referring now to Fig. 9, there is shown a video footage 500 incorporating a video zoom sequence 501 for which relative zoom ratio 502 are represented below each frame 503. The relative zoom ratio 501 is determined using motion vectors detected for each frame. Without departing from the scope or spirit of the invention, relative zoom ratios 502 can be computed, for example, from metadata encoded and associated with a video footage.

The encoded metadata may be obtained and recorded from a video camera, for instance, adapted with zoom sensors and circuitry capable of providing metadata on a state of a current zoom.

Turning to Fig. 10, there is shown a pan shot 600 comprising a video pan sequence of a cctage 601. In substantially the same manner as described

with reference to the zoom sequence 501, the pan shot can be detected from motion vectors. Optionally a panning sensor on a video camera can provide movement information of the video camera that can be used to determine an amount of panning in a pan-shot.

In the detection of zooms and pan-shots from a video footage, a predetermined zoom threshold level and a pan threshold level is optionally set which allows the detection of a zooms only where a relative zoom ratio, within a predetermined number of frames, is above the zoom threshold level and detected pan shots where a relative pan movement, within a predetermined number of frames, exceeds the pan threshold level respectively. For example, a zoom threshold level set at 3 with a predetermined number of frames set at 2, then a relative zoom ratio greater than 1:3 occurring within 2 frames would result in a detection of a zoom sequence.

In each case where a pan shot or a zoom is detected, a user is optionally given a choice to accept, or reject, for marking the detect pan or zoom accordingly. However, the process of marking a video footage so as to indicate locations of a pan shot or zoom can alternately be automated so that each zoom or pan detected can be marked based on predetermined criteria.

#### Printed Summary of Video Footage

The marking of pan shots and/or zooms is to readily identify those segments of video footage to which a special effect is to be applied or those segments of the video footage that are to be retrieved for printing, on a print device, or to generate icons for the footage. For example, the pan shot 600 of Fig. 10 is stitched together, that is, composited frame by frame to build up a still image pan as illustrated by Fig. 11. This process of stitching together video frames is known to those in the relevant art and can be done, for example, by determining a translation of predetermined picture elements, in a digitised representation, of each frame of the video footage.

Referring now to Fig. 12, there is shown a compositing of frames 701-703 of a video footage to produce a salient still image 704. Preferably the frames 701-703 to be composited into a salient still image, for printing on a print device, are extracted frames of a zoom portion of a video footage. Illustrated in Fig. 12 are three frames 701-703 selected from the zoom sequence shown in Fig. 9 and which are to be used to create the salient still image 704. To composite the salient still image 704, a first image represented by a frame 701 is scaled by a first predetermined factor and the scaled image used as a first portion 705 of the

salient still 704. In Fig. 12, the first portion 705 refers to the  $8 \times 8$  components as seen. Next, a second image represented by frame 702 is scaled by a second predetermined factor and the scaled second image is overlaid, or inserted, upon the first portion 705 to provide a second portion 706 of the salient still 704. Finally, a third image represented by frame 703 is overlaid (or inserted) upon the second portion 706 to provide a third portion 707 of the salient still 704. The resultant salient still image 704 comprise three portions 705-706 of the image 704 derive from different resolutions, composited together to form a single salient still image 704 which can be printed on a print device.

It is assumed that in a video footage where pan shots and zoom sequences have been detected, the photographer or cameraman taking the video footage had a desire to pan or zoom a particular scene and therefore there is a likelihood that the pan shot and zoom sequence of the video footage is of particular interest. Therefore images obtained from these particular scene are chosen for printing to a printer device as described with reference to Fig. 11 and Fig. 12.

Not all video footage includes pan shots and or zoom sequences and sometimes a pan or zoom does not contain images of particular interest. In these cases, images of the video footage intended for printing on a print device are selected based on motion analysis, or other characteristics of the video footage such as focus or slow motion blur. For example, determining local minima of motion occurring in select key frames is a good estimate that the cameraman has stopped (or paused) at an interesting point in a scene, or that characters pause on gestures to emphasise their importance, and therefore the selected key frames are of interest.

#### Motion Selected Transition

Referring to Fig. 13 there is illustrated a display device 900, displaying a first video segment 901 and at some time latter displaying a second video segment 902. The first displayed video segment 901 comprises motion occurring at a bottom right hand corner of the display 900. The motion is represented at the bottom right hand corner of the display 900 by a first set of motion vectors 903. A second set of motion vectors 904 for the second video segment 902 displayed on the display device 900 indicates that motion between the first 901 and second 902 video segments have changed from the bottom right hand corner of the display 900 to a left hand portion of the display 900.

A transition thus needs to occur if the first 901 and second 902 video segments are displayed consecutively. In accordance with the preferred embodiment, a transition type is chosen according to a location and an amount of motion on video segments to be displayed. Transition types, between a first video segment and a second video segment include: wipes which can be vertical, horizontal and or oblique; fades; and alike transitions.

Turning now to Fig. 14, there is shown an example of horizontal wipe 905 transition for the video segments of Fig. 13. The first video segment 901 is the outgoing scene while the second video segment 902 is the incoming scene of the transition 905. A horizontal wipe transition from right to left, on the display device, is automatically suggested (or chosen) because of the location of the motion in their respective video segments 901,902. The horizontal wipe transition 905 from right to left effectively transfers a viewers attention from the motion occurring in the first video segment to the motion occurring in the second video segment 902 and maintains the viewers attention while the transition occurs, slowly, across the display 900. The type of transition is chosen on the basis of a location and an amount of (visual) motion occurring between video segments. Those skilled in the art will recognised that a multitude of types of transitions can be devised, and corresponding motion rules associated to select a transition type, in place of the example described above, without departing from the spirit and scope of the present invention. For example, if visual characteristics are such that two connective video segments (clips) comprise a majority of a motion, and each segment is centrally localised, then a possible transition is that of fading-out a first video segment while fading-in a second video segment.

#### Colour Content

Another automated editing process based on a content of an image, in accordance with the preferred embodiment, is a selection of colours. For example, a selection of colour(s) for an opening title of a video production can, if a user so desires, be selected on the bases of statistical colour content information about an introductory portion of a video segment or on an entire video footage. Colours for titles, for example, can be selected from a complementary set of colours or a supplementary set of colours. The complementary and supplementary sets of colours are preferably chosen for a plurality of colour available in the video footage.

Referring now to Fig. 15, there is shown an image 1600 with text 1601 TITLE overlaid upon the image 1600. The image 1600 may represent a

printed image, a still image displayed on a display device or a predetermined portion of a video footage rendered and displayed on a display device.

Also shown in Fig. 15 are two sets of colours, those being a set of complementary colours 1602 and a set of supplementary colours 1603, which are to be used to provide a colour palette for the text 1601. The complementary 1602 and/or supplementary 1603 colours are selected based on colour content information of the image 1600. There are known techniques for determining dominant colours of an image for example, in a simple case an averaging of colour over predetermined regions of the image can provide a determination of dominant colours.

Referring to Fig. 16, there is shown a display device 400, with various portions 401 to 405 of the display 400 where a colour content of an image may or may not be of interest. For example, if graphical elements are to be used in a manner analogous to headers or footers of a document, the portions of an image where colour content information would be extracted is indicated by regions 401 and 402. Similarly, if text information (or graphical elements) is to be displayed at the left portion 403 or the right portion 404 of an image then colour content information is preferably extracted from these respective portions 403 and 404. Finally, if text information is to be centrally located, a centre portion 405 of the image is the portion from which colour content information is extracted. Colour content information is also used, in accordance with the preferred embodiment of the present invention, to guide a selection of a type of transition between video segments as previously described. Motion and/or colour content information can be used to select the type of transitions between video segments. An example of use of colour content information for selecting a transition type is, selecting a dissolve transition between a first video segment and a next video segment, where an average colour or alternatively luminance content of the first video segment is determined to be substantially similar to an average colour content of the next video segment.

The embodiment of the invention can also preferably be practised using a conventional general-purpose computer, such as the computer system 1700 shown in Fig. 17, wherein the processes described with reference to Fig. 2 to Fig. 6 are implemented as software executed on the computer system 1700. The computer system 1700 includes a computer module 1701, input devices 1702, 1703 and a display device 1704. The computer module 1701 comprises at least one processor unit 1705, a memory unit 1706 which typically formed of

semiconductor random access memory (RAM) and read only memory (ROM), input/output (I/O) interfaces including a video interface 1707, keyboard and mouse (or joystick) interface 1708. The storage device 1709 can include one or more of the following devices: a floppy disk, a hard disk drive, a CD-ROM drive or similar a non-volatile storage device known to those skilled in the art. The components 1705 to 1709 of the computer 1701, typically communicate via an interconnected bus and in a manner which results in a usual mode of operation of the compute system 1700 known to those in the relevant art. Examples of computers on which the embodiments can be practised include IBM-PC's and compatibles, Sun Sparcstations or alike computer systems evolved therefrom.

#### Enhanced Gathering Apparatus

Referring to Fig. 18, there is illustrated a block diagram overview of an Enhanced Gathering device 2100, comprising a content collection unit 2101, a processing unit 2104, a memory storage unit 2103, a user interface unit 2105, a read only memory (ROM) 2108, a communications port 2106 and an output interface 2102, which is an example of an information gathering device.

Optionally, and depending upon a predetermined application of the enhanced gathering (capture) device 2100, the communications port can be removed and the memory unit 2103 replaced with technology such as PCMCIA, bubble memory, a floppy disk or any portable memory device.

Generated (or gathered) output information can be communicated to external devices via the output interface 2102 and represents content gathered. The content gathered information can be used as input information to external devices. For example, external devices can include content recording devices such as analogue or digital video records, optical disk drives, sound recording devices, image recording devices, printer devices or general purpose computers (eg. IBM compatible PC, Apple Compatible PC, Sun Workstation) to name a few. The content gathered information includes content data, and preferably metadata associated with the content data. An example of content data is image data gathered by a video camera device and the associated metadata can include such information as the date, time, camera settings and location where the image data was taken.

The Content Collecting Unit (CCU) 2101 preferably comprises a plurality of content sensors, each capable of receiving (collecting or gathering) predetermined forms of content information. Examples of such sensors include a microphone for receiving sound (content) data; a charge coupled device (CCD)

for collecting image data; and a light sensor for determining and collecting lighting conditions (metadata) associated with collecting image data. Preferably, the CCU 2101 further includes elements (transducers) that gather sensory information about a current state of various content sensors, or associated hardware such as the current zoom factor on a video camera. This additional data is typically available for use by the processing unit 104 and can be encoded as part of the metadata in the gathered content information.

At a functionality level, a user selects at least one of a plurality of optional templates which are made available to the user through the user interface unit 2105. The user controls the enhanced gathering device 2100 through a series of selecting buttons (not illustrated) to make a desired template selection. The processing unit 2104, in communication with the memory storage unit 2103 and/or the ROM 2108, provides, via the user interface unit 2105, the various optional templates currently available on the enhanced gathering device 2100.

Each template typically comprises a series of template instructions which can be described by a predetermined programming language. For example, the programming language may be specific to an application and proprietary to the device 2100 or can be an implementation of a general purpose programming language such as Java (Trade Mark of Sun Microsystems Inc). Template instructions are uploaded, via the communications port 2106, from a general-purpose computer system 2107 to the enhanced gathering device 2100 and are stored in the memory storage unit 2103.

Alternatively, cartridge systems which are commonly used for computer games and/or flashcards for storing the templates can be used as a plug in device to upload templates to the enhanced gathering device 2100. Preferably, part of the series of template instructions, in particular, instructions which are common to several templates can be stored in Read Only Memory 2108 to reduce the amount of (RAM) memory required in the storage unit 2103. The template instructions are then available to the processing unit 2104 for runtime calculations. Preferably, the general purpose computer system 2107 illustrated in Fig. 18 is also used as the external device previously described which interfaces with output interface 2102 to download the content information.

A user interface device 2105, for instance, comprising a liquid crystal display (LCD) and a series of control buttons, is used to convey (messages) information to the user, as well as gathering user input in response to those messages.



The processing unit 2104 interprets the template instructions to provide options including content analysis of the content information, presenting to the user various choices for selection of menus available and interpreting a response to the menus from the user. The choices selected by the user, influence or change: the current state of the CCU 2101; user preference settings; and/or the current template being used by the device 2100. Preferably, some of these tasks are fully automated in that the template makes the best decision from the content information available and adjusts the CCU accordingly. For example, in a first embodiment of the enhanced gathering apparatus described hereinafter, the template may be used to decide from predetermined conditions that a portrait view of an image is desired over a landscape view and therefore a cropping of the image to a portrait view is performed if necessary.

Fig. 28 is a flow diagram illustrating the general operation of the enhancing gathered information. The process commences at step 21100. In step 21102, at least one template is uploaded in the enhanced gathering device. In step 21104, at least one form of content information is received in the enhanced gathering device. In step 21106, template instructions of the uploaded template are executed to generate composition instructions based on the received content information. In step 21108, a composition of the content information is enhanced in accordance with the composition instructions. In step 21110, processing terminates. Preferably, the method further involves storing the enhanced content information on a storage means and may optionally be downloaded to an external device.

#### First Enhanced Gathering Apparatus

The first enhanced gathering apparatus is now described, by way of example, with reference to a Digital Still Camera where preferably the enhance capture (gathering) device 2100 is inbuilt or forms part of an enhancement system of the camera. In this example the content collecting unit 2101 comprises an optical lens system and CCD as is typical for such cameras and can further include distance measuring sensors, light intensity determination devices, focus detection devices and position locating devices (eg. Global Positioning System GPS). These sensors and devices gather image data, and metadata such as a location where the photograph was taken. The user interface unit 2105 for the Digital Still Camera is a liquid crystal display (LCD) and preferably the user interface includes devices for entering data (eg. buttons). The communication

port 2106 is used to upload one or more required templates, from a general purpose computer 2107, to be stored in the memory storage 2103.

Fig. 19 illustrates a schematic representation of a digital still camera 2202, used in taking a picture of the subject matter 2201. As a part of the digital still camera 2202 there is a LCD 2203, in which the image to be recorded is displayed. A user can elect to record the image by pressing shutter button 2206 or can enter a template mode by pressing a mode template button 2204. The user is allowed to exit the template mode by pressing the mode template button 2204 a second time.

Turning now to Fig. 20, upon pressing the mode template button 2204 a list of available templates 2301 is displayed on the LCD 2203. In this example, the templates include a landscape template, a pan shot template and a portrait template, having been previously uploaded as described above. A selector button 2205 is used to navigate through the available selections, and select the desired template. A selection arrow 2302 displayed on the LCD 2203 indicates a current template mode selection.

Once a selection of template is made, for example selecting the Portrait template, the image displayed on the LCD is analysed by the processing unit 2104. The results of the analysis are used by the template to guide the user. In this example, the portrait template requires that for best results the subject matter 2201 should be centred and occupy at least a predetermined percentage of the visible area of the display (ie. image to be recorded).

Referring now to Fig. 21, there is shown the LCD 2203 after the Portrait template has been selected. A number of elements appear on the liquid crystal display 2203 including a positioning indicator 2401, a line up marker 2404 and a centre marker 2405 which would not appear on a photograph (or recorded image), should the user desire to take a record of the image by pressing the shutter button 2206. Preferably, details of the current status of the positioning indicator 2401, the line up marker 2404 and the centre marker 2405 are generated and included as metadata, and associated with the corresponding image, for later use or evaluation. For example, if the user has not correctly aligned the relevant markers 2404, 2405 and /or positioning indicator 2401, and the content collecting unit is capable of gathering a larger view of an image than is desired, then the information part of the metadata, relating to the markers 2404, 2405 and the information of the positioning indicator 2401, can be used to evaluate a preferred print area for the image. The image can then be cropped to a predetermined print area.

Preferably, the image and corresponding metadata is first downloaded to a general purpose computer which is used to edit the image, if necessary, before cropping the image to the predetermined print area and outputting to the output device (eg. printer device).

The positioning indicator 2401 can advise the user to use another orientation, such as a vertical orientation of the camera 2202, by flashing on and off a peripheral arrow 2402 on the indicator. The position indicator further provides zoom advice to the user. A "zoom in" arrow 2403 indicates that it is desirable to zoom in on the subject 2201 of the image. Accordingly, a "zoom out" arrow 2406 can flash to indicate that it is desirable to zoom out. Alternately, the subject 2201 or the camera 2202 can be moved closer together or farther apart to achieve a substantially similar effect as a zoom in or zoom out respectively.

The centre marker 2405 and the line up marker 2404 are provided for a variation on the first enhanced gathering apparatus in which the enhanced capture device 2100 is capable of analysing the content information of the image and substantially detecting an approximate position of the subject 2201 in the image.

The detection of the approximate position of the subject of the image need not be performed by analysing the content information and can be determined by other techniques. For example, the position of the subject in the image can be determined by detecting which portion of the image a user is looking at from a relative position of the user's eye pupils in relation to the image being observed.

In this variation of the first enhanced gathering apparatus, the centre marker 2405 is placed at the centre of the visible area, while the line up marker 2404 is placed at the centre of a detected subject. Since the line up marker attempts to follow the subject 2201, a user can manipulate the line up marker 2404 and centre marker 2405 by moving the camera 2202 to align the markers.

As illustrated in Fig. 22, the user preferably rotates the camera 2202 into the desired vertical orientation so that the peripheral arrow 2402 no longer flashes on, indicating a correct orientation. At this stage, as illustrated in Fig. 22, the markers 2404, 2405 do not necessarily align, the subject 2201 does not occupy a predetermined portion of the image and the "zoom in" arrow 2403 is still indicating that a zoom in is desired.

Referring now to Fig. 23, the user has centred and zoomed in on the subject matter 2201 of the image correctly, which results in all the arrows on the position indicator being off and preferably an indicator 2602 of alignment also

appears on the LCD 2203. By not indicating any further desired changes to the zoom or orientation, the position indicator is suggestive that the composition of the image (photograph), in accordance with the chosen template, is correct in zoom and orientation and that the current image (photograph) is recommended for recording. The user activates the shutter button 2205 to record the image.

Referring to Fig. 24, there is shown a further example of the Digital Still Camera operating in portrait template mode where the subject of the image comprises more than one object. The template attempts to optimise a best fit of the objects onto a visual area of the image. Preferably, a geometric centre of a bounding box containing the objects is used as an align marker 2702 location and a geometric centre of the visual area of the image (ie. visible display area 2203) as a centre marker 2701. Further, since one of the objects appears partly outside the visible display area 2203, the position indicator recommends, by flashing on the zoom out arrow 2406, that a zoom out is desirable.

A substantially similar description applies to the landscape template and the pan shot template where the landscape template is generally used for photographing landscape scenery and the pan shot for panoramic wide view shots. For the sake of brevity and clarity, the templates have been described as providing a small number of features such as orientation, zoom, etc. in which the template can provide advice to a user. However, the template is not limited thereto. For example, one or more templates could provide the user with advice on features including desired lighting for a photograph, best angle for lighting in a photographic studio type setting, and camera focus advice to improve composition of a photograph. The camera focus advice can be by way of suggesting a camera setting to achieve a background focus and foreground defocus. The templates attempt to provide advice to the user, from content information available to the templates, which could otherwise be provided by one skilled in the art of photography. For example, advice on preferred orientation, best zoom and positioning of subject matter to provide guidance to the amateur photographer who is typically not abreast of the required skills to produce a quality photograph or a photograph of a predetermined style.

Templates are implemented in code with the use of a programming language as previously described. Pseudocode for a portrait template is included in the accompanying Appendix A. The main routine of the pseudocode executes and runs while the digital still camera is powered up and analyses an image in the

viewfinder (LCD 2203) to check for various conditions that are used to detect which of the available template features are optimum.

When the user has pressed the template mode selection button 2204 the available templates are displayed. The user can select a desired template, with the selector button 2205 and then exit the template selection mode. Upon making a selection, a software routine, such as that represented by the pseudocode contained in Appendix A, checks for various conditions that are required for zoom and orientation of the image in the viewfinder.

The sub routine SetUpForTemplate returns a TRUE Boolean value when the image is aligned and the main program "Void Main()" will flag that the image (picture) is aligned.

As previously described, additional information such as metadata can be recorded and associated with an image. Examples of this additional information include date, time, actual zoom used, shutter speed, aperture, that the still is a portrait of a person, the number of people in the shot and the centre of each face detected in the shot as well as the geometric centre of all the faces detected.

#### Second Enhanced Gathering Apparatus

As described hereinafter, a second enhanced gathering is substantially the same as the first except that the enhanced capture system 2104 forms part of a video camera. Templates are provided capable of providing advice on features including zoom, pan, composition and directing (in the sense of a film direction) of a video clip. For example, a plurality of uploadable templates can be provided to a user, for uploading into the video camera, including templates for directing a wedding scene; a real estate agent's home selling promotional clip; and a country holiday scene which may include suggestions as to sights and wonders of the world that are recommended for video recording. For instance, the user may be going on a holiday to Egypt and taking a video camera comprising an enhanced capture system 2104. To enhance the composition of a video recording, which will be taken while the user is in Egypt, a template of preferred sights of North Africa can be uploaded to the enhanced capture system 2104. Besides performing the other function herein described, the template may further suggest recommended sights and wonders worth video recording (eg. the Pyramids).

As a further example, reference is made to an uploadable template for directing a wedding scene (hereinafter a wedding template). A user desires to make a video recording of a relative's wedding; the person calls up on a general purpose computer a wedding template and uploads it to the video camera,

comprising the enhanced capture system 2104, via the communications port 2106. The wedding template is stored as (compiled) code in the memory storage means 2103 while generic features to uploadable templates have been previously stored in ROM 2108 at, for instance, manufacture of the video camera-enhanced capture device. The generic features may include details of maximum and minimum zoom settings available, aperture settings possible and other features that may be common to a majority of templates for the video camera.

When the user starts to record a wedding scene using the video recorder, the wedding template is activated and advises the user as to the shots the person must take in order to produce a video clip, which upon editing if necessary, complies to a substantially professional video recording. For example, the user is prompted, preferably in chronological order, with shots that provide a desired wedding scene composition. That is, the user is prompted to take a shot of the groom entering the church, the groom standing at the altar, the bride leaving the car and making her way up the steps of the church, the bride walking down the aisle and so on. Further, the user is also advised on technical video issues such as zooming, panning, lighting and audio. For example, if when taking the shot of the bride walking down the aisle, the user zooms too fast on the subject, a "slow down" indicator is flashed in the viewfinder of the video camera and a too slow zoom is equally signalled to speed up the zooming.

Preferably, a user indicates through a series of menu driven commands displayable on a viewfinder (LCD) which section, or portion the user is currently video recording. The user ticks off (selects) the section that is to be, or has just been, video recorded. For example, in the wedding template, the user is given a display of a multitude of sections, or portions, which is recommended for videcing including the aisle section, when the bride is walking down the aisle. The user ticks off the aisle section before or after recording the relevant scene. Data associated with the "ticking off" of a section is recorded as metadata for the corresponding portion of the video clip. For example, information that a portion (section) of the wedding scene is the section where the bride walks down the aisle is recorded as part of the metadata to the wedding scene. At editing of the recorded wedding scene, preferably by an external device such as a computer adapted to edit video clips, the metadata associated with the wedding scene can be used to identify various sections of a wedding scene.

Third Enhanced Gathering Apparatus

A third enhanced gathering apparatus is described with reference to a sound mixing desk comprising enhanced capture device. A sound mixing desk is typically used for gathering musical content from a number of different channels, applying various effects and filters to individual and/or groups of channels, then combining these channels into one or more output channels that are then recorded, or used to drive amplifiers, for an audience.

Again for the sake of simplicity, a simple two-channel interface sound mixing desk is described. The mixing desk is implemented using a computer to display controls, analyse the content, with the aid of filters, analysers, effects units and alike hardware assistance, provide a user feedback and send control signals to the filters and effects units.

Referring to Fig. 25, there is shown a sound mixing desk system 2800 comprising: a plurality content gathering sensors 2801 such as sound transducers; an enhanced gathering device 2802 capable of supporting multiple channels; a general-purpose computer system 2804; a plurality of effects units, filter banks and analysers 2810; an output amplifier 2809 with at least one output port 2811; and playback devices 2812 including compact disk CD player, reel to reel tape and/or cassette tape unit.

Information gathered by the sensors 2801, and the other peripheral devices 2812 is passed to the enhanced capture device 2802 which processes the gathered information according to a selected template. A desired collection of templates is uploaded to the enhanced gathering device 2802 by a data (omnibus) link 2805 to be stored in a memory storage unit of the device 2802. While the type of content gathering sensors may vary from implementation to implementation, the enhanced gathering device 2802 comprises substantially the same components as illustrated in Fig. 18. In the present example of a sound mixing desk 2800, a user interface unit of Fig. 18 is excluded, preferably, in favour of an external display 2813, keyboard 2814 and mouse 2806 device.

The information gathered by the sensors 2801 and other peripheral devices 2812 form part of the content information. The content information preferably includes metadata such as time, date, recording duration length, musical style and title of track, which may be added to the content information by a selected template. Although a plurality of templates are uploaded to the enhanced gathering device 2802, a single template selection can be made by a user through manipulation of a mouse and/or keyboard device driven menu displayed on the display device 2813. The information gathered is directed, via

a communication link 2803 between the enhanced gathering device 2802 and the plurality of effect units, filter banks and analysers 2810, to the output amplifier 2809 and output port 2811 for use. Content information is passed to the computer 2804 through the omnibus link 2805, and data about content information is displayed for the user on the display device 2813. Preferably, the content information is further analysed by a central processing unit which extracts data to be displayed. Alternatively, the content information is analysed by the enhanced gathering device and only relevant display data is transmitted via link 2805 to the computer 2804.

The user is allowed to interact with various settings through manipulation of the input devices 2806 and 2814. The various settings are conveyed on a bi-directional link 2807, as control signals, to the amplifier 2809 and to the plurality of effect units, filter banks and analysers 2810 that are designed to alter parameters of effects, filters to produce desired modifications to the information output on 2811. The bi-directional link 2807 preferably also carries signals generated by a number of devices, including the plurality of effects units, filter banks, analysers 2810, the output amplifier 2809 and the playback devices 2812, relating to a current status and/or playback content information of the various playback devices 2812 if required.

A user can alter preference and default settings through the computer system 2804. The default settings are generally dictated by the type and/or combination of components connected at a sound mixing desk. For instance, different choices of transducers or playback devices may require different default settings. Preference settings generally provide a manner of customising a sound mixing system to user preferred settings. Both preference and default settings are stored in non-volatile memory, generally part of the computer system 2804.

An example of an uploadable template is described below with reference to a band mix template. A band mix template provides advice to a user, of the sound mixing system, associated with sound mixing (and if desired recording) of musical bands, including such band styles as a String Quartet, a Rock and Roll band, a Heavy Metal band and a Classical Orchestra. At selection (and execution) of the band mix template, the user is prompted to tick off (select) a music style of the band, a size of the venue (ie small hall, outdoors, studio etc). Preferably, the user is further prompted to identify which channels (or associated sensors) are to be used to gather information (sound signals) for instruments and which sensors are for vocals. The band mix template then produces a



recommended list that provides best allocation of available microphones (sensors) for each channel.

Further, the user may select a test run ("dry run"), upon which the band mix template can analyse the content information and advise on which sensor is picking up reflections off surrounding objects, and either nominate a solution (eg. re-allocate the microphones or re-position the sensors) or attempt to electronically compensate for the distortion through the effects and filters available. Similarly, if the band does a dry run of a performance in the correct order, a sound level setting can be determined and used as a guide for an eventual performance.

As data (information) is collected by the sensors 2801, the enhanced gathering device 2802 performs content information extraction and provides this content information to the computer 2804. Preferably, from the derived content information, the computer 2804 sends compensation control signals, for the sensors on each channel, to adjust the filters (graphics equalisation), on each channel. The computer 2804 further sends control signals that adjust the effects and also fine tune the filters on each channel to all the instruments in a band, to conform with the music style of the template.

For example, a user can select a Heavy Metal music template with a high distortion level to the vocal channel. Boosting of the top end frequencies is recommended to the user of the mixing desk and preferably these features are adopted automatically unless the user desires otherwise.

However, if the user selects a Classical music template, then all the channels are biased to a sound level setting to be as flat as possible. That is, no (special) effects, through the effect units and/or filter banks, are added to the channels so that instrument sounds are reproduced as closely as possible to a true sound of the instrument.

As previously described the template provides advice to the user which could otherwise be provide by one skilled in the art. In the present example the art is the art of sound mixing, for example advice on echo, sound reflection, reverberation, tempo are at least some features that a sound engineer would advise on when recording and sound mixing.

Whilst only briefly described with reference to the second enhanced gathering apparatus, an optional preferred feature of the templates applicable to each of the above described apparatus is that of providing a check list of events or actions that a user can tick off (select) in respect of predetermined events or actions. For instance, in the digital still camera example described above, a

check list of predetermined shots is provided to a user through the user interface. Before (or after) the user takes a shot (photograph), the user ticks off a corresponding predetermined shot from the check list. For example, the digital still camera is uploaded with a birthday party template which includes a birthday party check list. The check list is displayed on the user interface and the check list may include, by way of example, the following:

- \* birthday party games shot;
- \* blowing out of the candles shot; and
- \* making a toast to the birthday person shot.

The user takes photographs of the birthday party and if the photograph taken coincides with a shot described on the check list - the user can tick off (mark or select) that a predetermined photograph described in the check list has been taken. Accordingly, at ticking off a selected shot from the check list, metadata is generated and associated with the photograph (or image). This metadata information can be utilised, for example, when the photograph (or image) is downloaded to a general purpose computer, to put together a photo-album of the birthday party.

A plurality of check lists can be provided with each template that can assist the user in determining which photographic shots would make good composition when putting a photographic album together.

In the example of the sound mixing desk, a substantially similar check list for music can be provided where a user ticks off the check list once an event (eg. musical piece, mixing characteristics) described in the list is performed. Again, corresponding metadata is generated and associated with the event. The corresponding metadata can be a simple identifier for the event or a more complex description of the event as may be desired for future manipulation or editing of the event.

Referring now to Fig. 26, there is shown another example of a check list displayed on the user interface display (eg. viewfinder 2203 of the digital still camera of Fig. 19 or the viewfinder of the video camera of the second enhanced gathering apparatus hereinbefore described). The check list 2900 is associated with a real estate template which can be uploaded (load) into the enhanced gather capture device 2100 as previously described with reference to the first and/or second enhanced gathering apparatus. The real estate template advises a user (typically a real estate salesperson), of the digital still camera or the video camera on composition and/or images required to compose a good presentation. The

real estate template can, as an optional function, provide a real estate check list 2900 comprising: description fields 2901 which provide a brief description of a recommended photograph (or video clip in the case of the video camera) to be recorded; check list indicator fields 2902 provided for marking off each corresponding image (photograph) recorded; and check list marker 2903 to indicate which of the recommended images have been recorded.

The check list marker 2903 is typically added by the user, through an available user input interface (eg. buttons), preferably when the recommended image has been recorded. The real estate template, upon checking off the items of the check list, generates metadata and associates the metadata with the corresponding image of the check list. Preferably, the metadata is an encoded description of the corresponding image to which it is associated. Further, the metadata can include such characteristics as time, date, location etc.

Referring to Fig. 27, there is illustrated a block diagram of information flow from an enhanced capture (gather) device 21000 to a computer system 21001 comprising uploadable templates 21002 and subsequently to an output device 21003. A plurality of templates 21002 are uploaded, via a bi-directional communication link 21005, typically from a computer system 21001 comprising a set of templates 21002.

Content information 21006 is gathered by the input devices 21000, and metadata 21007 is generated by at least one of the uploaded templates 21004. The generated metadata 21007 can be as simple as a description of the content information determined by a user response to a check list and/or determined from the template by analysis of the content information 21006. The generated metadata 21007 is associated by the template to a corresponding part or all of the content information 21006. For example, in the case of the real estate template - the content information may comprise an image from a recommended check-list 2900 of images suggested by the real estate template. The corresponding metadata generated for the image can be based upon features including: content analysis directed to orientation, average colour, average brightness, dominant colours, predetermined patterns, statistical analysis in spatial, time or frequency domains identified in the content information; user preferences including colour, arrangement of information; and device specific features including a current state of hardware components, predetermined formatting of information for downloading to various output devices.

The enhanced gathered information, which includes content information and metadata is downloaded to an external device. In the present example, the external device is a general purpose computer system 21001 as previously described.

However, the external device can be a dedicated external device capable of receiving the enhanced gathered information. Such dedicated devices include: a printer device, a display device (eg. cathode ray tube, CTR, Liquid Crystal Display (LCD)), a photocopying device, and/or an acoustic recording device, each of which have been adapted to receive the gathered information and provide an output or storage of the information according to predetermined preferences. The external device can be, for example, a video cassette tape for recording and storage of the enhanced gathered information to be used at a latter stage in composing a final output.

Preferably, an uploaded template 21004 formats and arranges the enhanced gathered information in accordance with predetermined setting or formats the enhanced gathered information depending upon the type of external device to which the enhanced gathered information is to be downloaded to.

The process of enhancing gathered information may be practiced as software, readable and executable by a processing device such as a microprocessor, recorded on a computer readable recording medium. The recording medium includes magnetic and optical storage memory such as a hard and/or floppy drive, magnetic tape, CD-ROM, magneto-optical disk, a memory card, read-only memory, etc. Further the software may be resident on another computer system accessed through a Network system or across the Internet.

Examples of enhanced gathering devices that may be used in implementations of the present invention include the following digital video camcorder devices:

SONY VX-1000 (Sony Corporation, Japan);

SONY PC-10 (Sony Corporation, Japan);

PANASONIC NV-DS5 (Matsushita Electric Corp, Japan);

CANON MV1 (Canon Inc., Japan); and

CANON XL1 (Canon Inc, Japan).

These devices provide for the generation of a limited set of metadata relating to recorded audiovisual information. Such metadata typically relates to rudimentary identification of the information. More advance devices include the:

KODAK DC220, 260 (Kodak Corp., USA); and  
 MINOLTA Dimage EX Zoom 1500 (Minolta Corp., Japan),  
 both of which are still image digital cameras and may generate higher level  
 metadata, for example corresponding to the "real estate" embodiment described  
 with reference to Fig. 26.

The foregoing describes only some embodiments of the present invention,  
 however, modifications and/or changes can be made thereto without departing  
 from the scope and spirit of the invention.

#### APPENDIX A

##### Pseudo Code

// Portrait Template

Boolean SetUpForPortrait(Boolean UI\_Flags[], image)

```
{
    Boolean Result = TRUE
    // Assume only called IF 0 < numFacesDetected < 4
    IF InadequateMarginAroundGeometricCentre(image) THEN
        UI_Flags[ZOOM_OUT] = TRUE
        Result = FALSE
    ELSEIF TooMuchMarginAroundGeometricCentre(image) THEN
        UI_Flags[ZOOM_IN] = TRUE
        Result = FALSE
    ENDIF
    IF ((FacesBoundingBoxTallerThanItIsWide(image) AND
        CameraOrientation Equals LANDSCAPE) OR
        (FacesBoundingBoxWidererThanItIsTall(image) AND
        CameraOrientation Equals PORTRAIT)) Then
        UI_Flags[CHANGE_ORIENTATION] = TRUE
        Result = FALSE
    ENDIF
    IF (FacesBoundingBoxCentred(image) EQUALS FALSE THEN
        UI_FLAGS[CENTRE_BOUNDINGBOX] = TRUE
        Result = FALSE
    ENDIF
}
```

```
RETURN Result
}
```

```
Void main()
```

```
Boolean UIFlags(CONST_NUM_UI_FLAGS)
```

```
ImageType image
```

```
TemplateType templateSelected
```

```
WHILE PoweredOn
```

```
IF MenuButtonPressed() Then
```

```
    templateSelected = DisplayAvailableTemplates()
```

```
ENDIF
```

```
// Get the image that the camera is looking at
```

```
GetImageInViewfinder(image)
```

```
// Check for cases for which there is a template
```

```
IF templateSelected EQUALS CONST_PORTRAT_TEMPLATE
```

```
THEN
```

```
    IF SetUpForPortrait(UIFlags, Image) EQUALS TRUE
```

```
    THEN
```

```
        UIFlags[PICTURE_ALIGNED] =
```

```
        TRUE
```

```
    ENDIF
```

```
    ELSEIF .....
```

```
        .... Other checks for other templates
```

```
    ENDIF
```

```
    UpdateUI(UIFlags)
```

```
END WHILE
```

```
}
```

#### 4 Brief Description of Drawings

Notwithstanding any other forms which may fall within the scope of the present invention, preferred forms of the invention will now be described, by way of example only, with reference to the accompanying drawings, in which:

Fig. 1 is a schematic representation of an overview of an apparatus in accordance with a preferred embodiment of the present invention;

Fig. 2 is a block diagram of a sample input script of multimedia elements generated in accordance with the preferred embodiment;

Fig. 3 is a flow chart of a beat synchronisation process in accordance with the preferred embodiment;

Fig. 4 is a flow chart of a transition generating process in accordance with the preferred embodiment;

Fig. 5 is a flow chart of an overview of an editing process in accordance with the preferred embodiment;

Fig. 6 is a flow chart a portion of editing process of Fig. 5 in more detail;

Fig. 7 illustrates an example of beat synchronisation between multimedia elements in accordance with the preferred embodiment;

Fig. 8 illustrates an example of audio level editing in accordance with the preferred embodiment;

Fig. 9 illustrates an example of a zoom sequence forming part of a visual element;

Fig. 10 illustrates an example of a pan sequence forming part of a visual element;

Fig. 11 represents an image derived from stitching together the pan sequence of Fig. 10;

Fig. 12 represents a salient still image derived from a compositing of the zoom sequence of Fig. 9;

Fig. 13 represents motion vectors of two video sequences indicating a motion when the video sequences are rendered to a display device;

Fig. 14 represents an example of a transition between video sequences (segments);

Fig. 15 shows an example of an image and overlay graphics TITLE wherein a colour of the overlay graphics is chosen in accordance with the preferred embodiment;

Fig. 16 illustrates an example of likely areas of interest for extracting colour content information from an image;

Fig. 17 is a schematic representation of a computer system used to implement the preferred embodiment;

Fig. 18 is a block diagram representing an overview of an enhanced gathering device;

Fig. 19 is a representation of Digital Still Camera, comprising the enhanced gathering device of Fig. 18;

Fig. 20 illustrates a "Template selection" mode of the Digital Still Camera of Fig. 19;

Fig. 21 illustrates composition features, "Orientation Guidance" and "Alignment", of the Digital Still Camera of Fig. 19;

Fig. 22 illustrates another composition feature, "Zoom Guidance", of the Digital Still Camera of Fig. 19;

Fig. 23 illustrates an example of an enhanced image composition;

Fig. 24 is an example of a viewfinder display for a portrait of more than one person in the image;

Fig. 25 is a block diagram of an alternate enhanced gathering device;

Fig. 26 is an example of a check list associated with a template;

Fig. 27 is a block diagram of information flow for the enhanced gathering devices; and

Fig. 28 is a flow diagram of the process of enhancing gathered information.



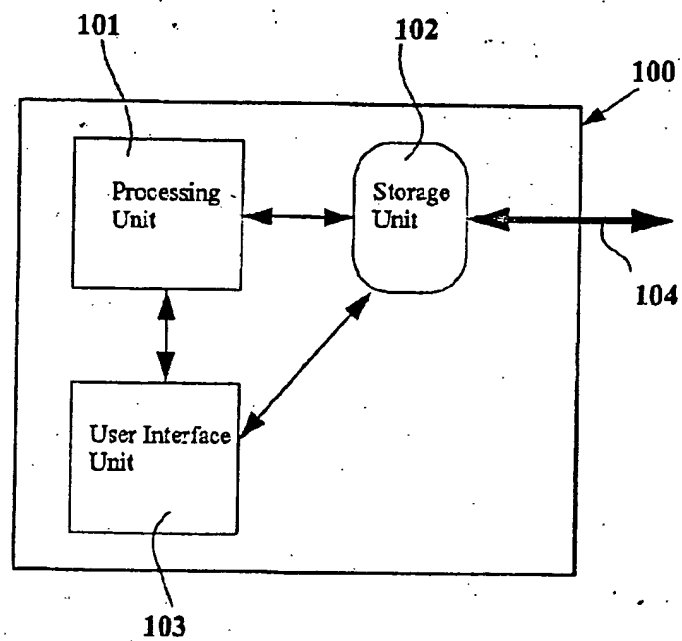


Fig. 1.

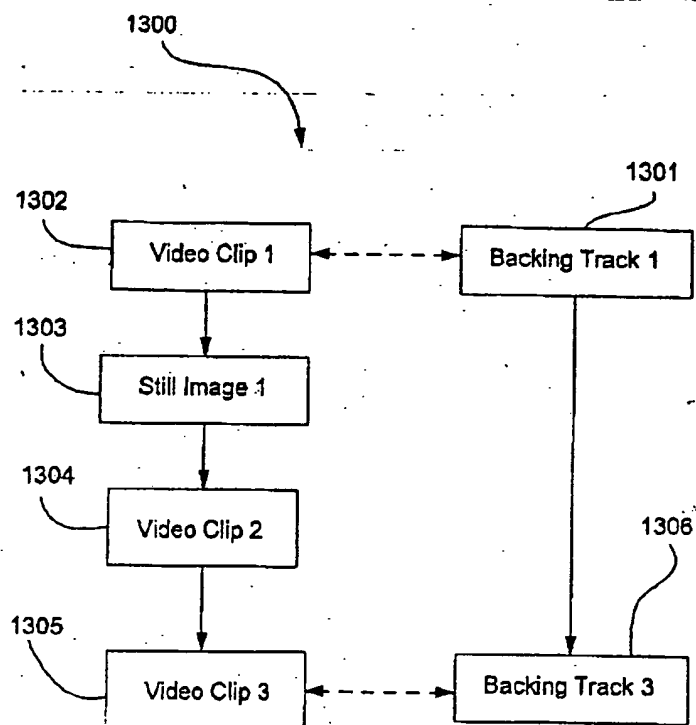


FIG. 2

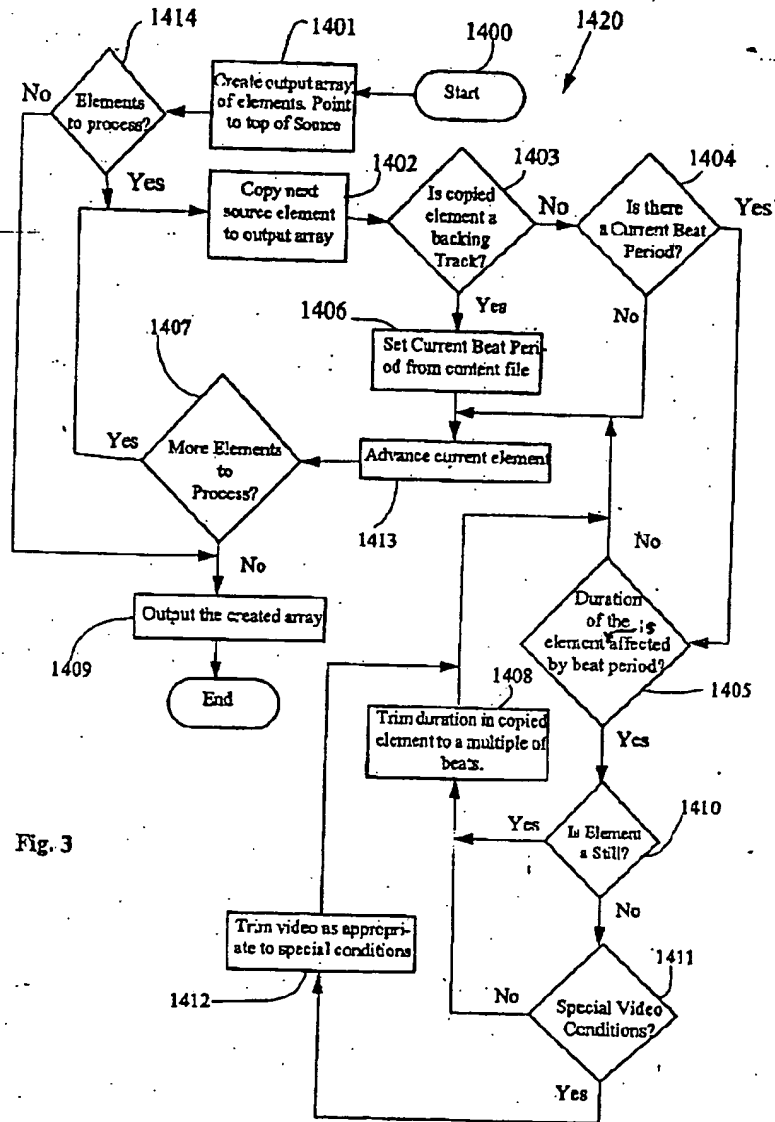


Fig. 3

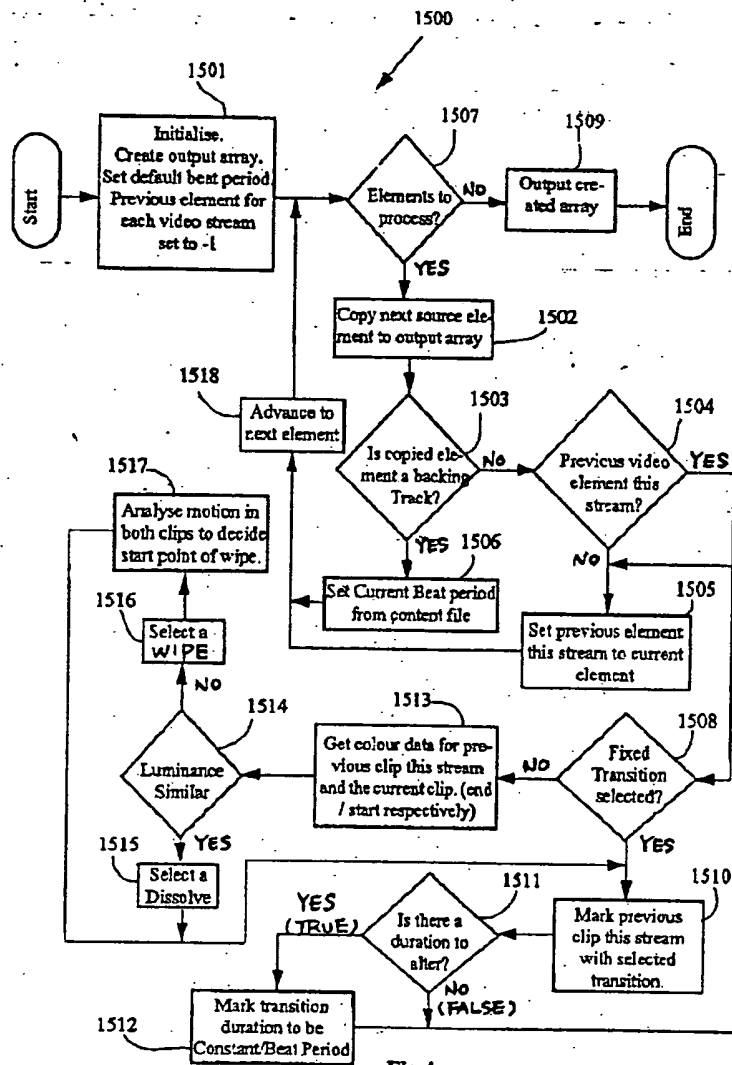


Fig 4

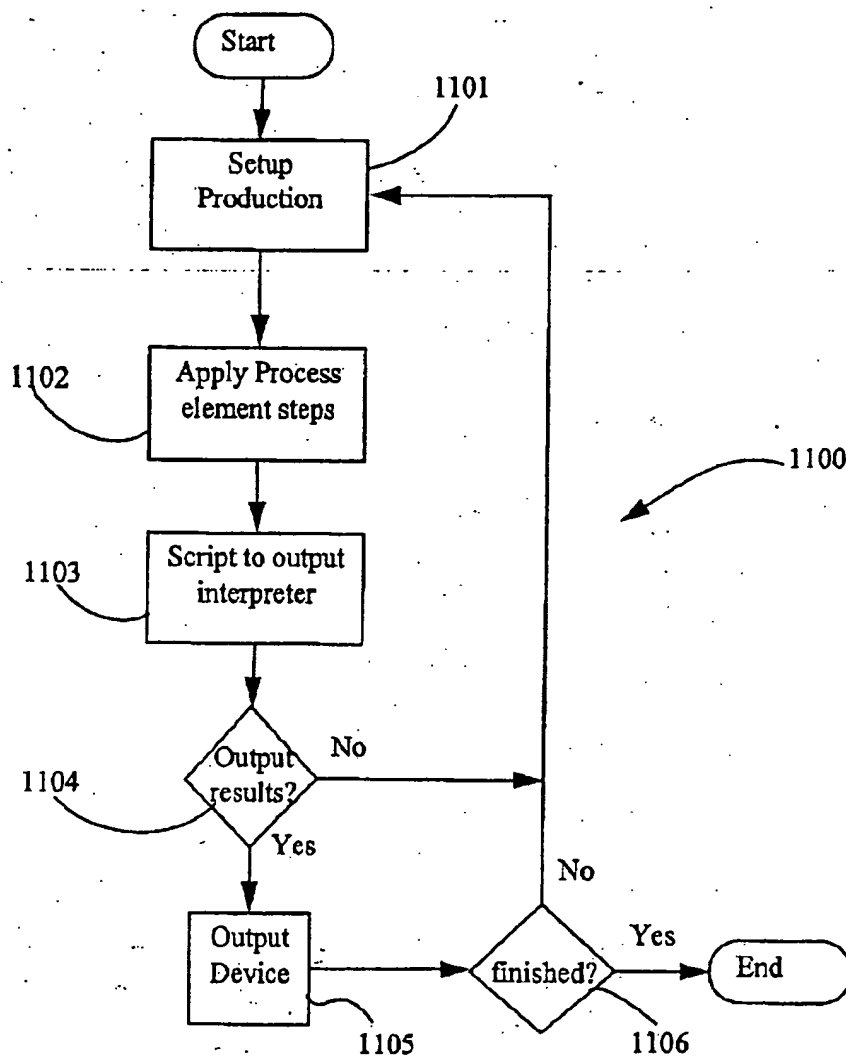


Fig. 5

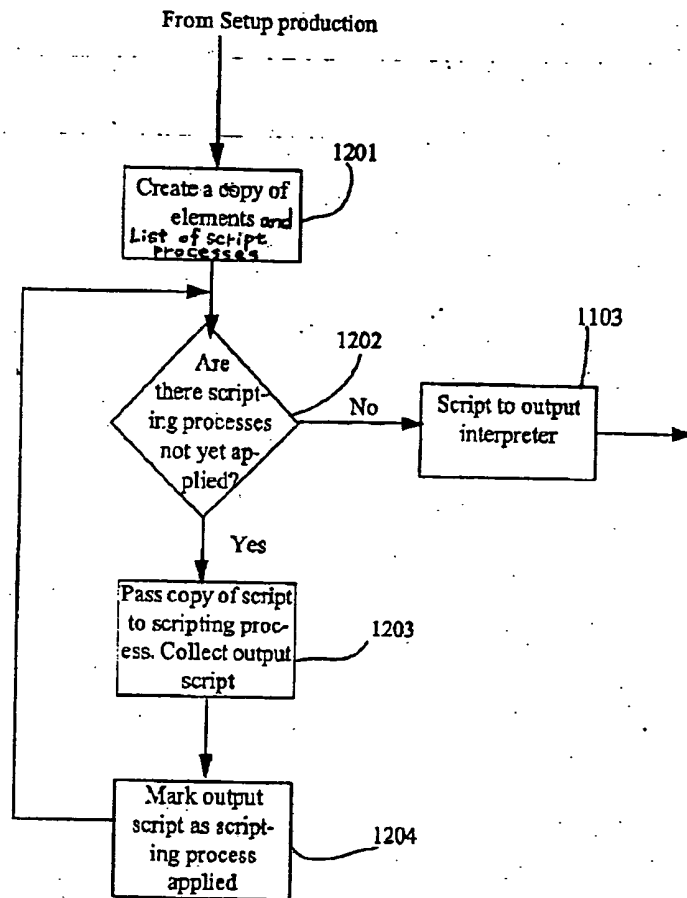


Fig. 6

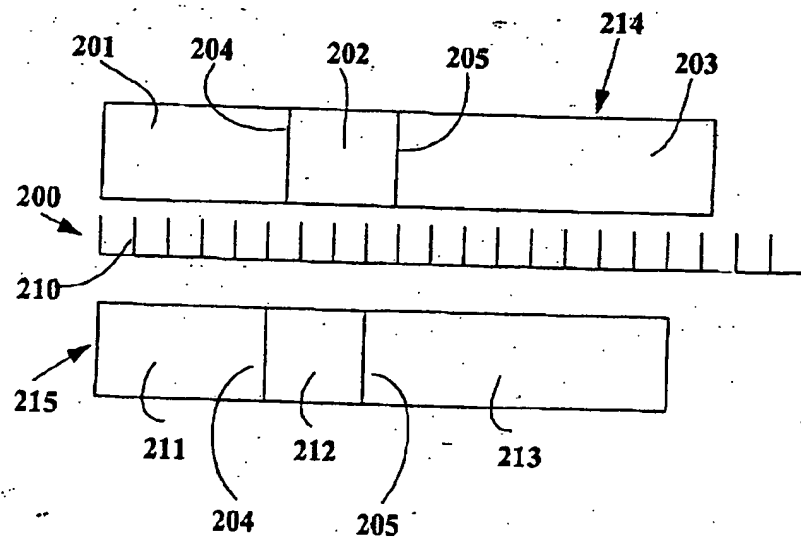


Fig. 7





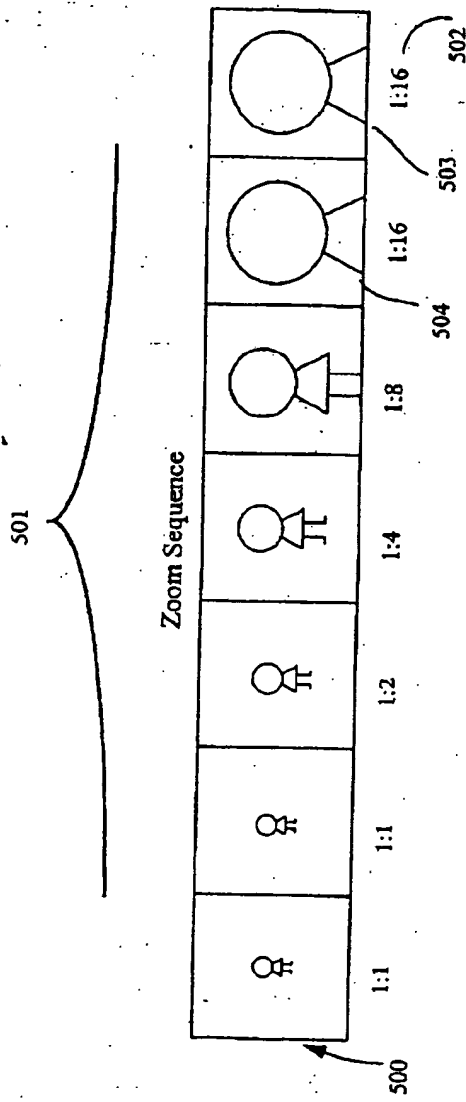


Fig. 9

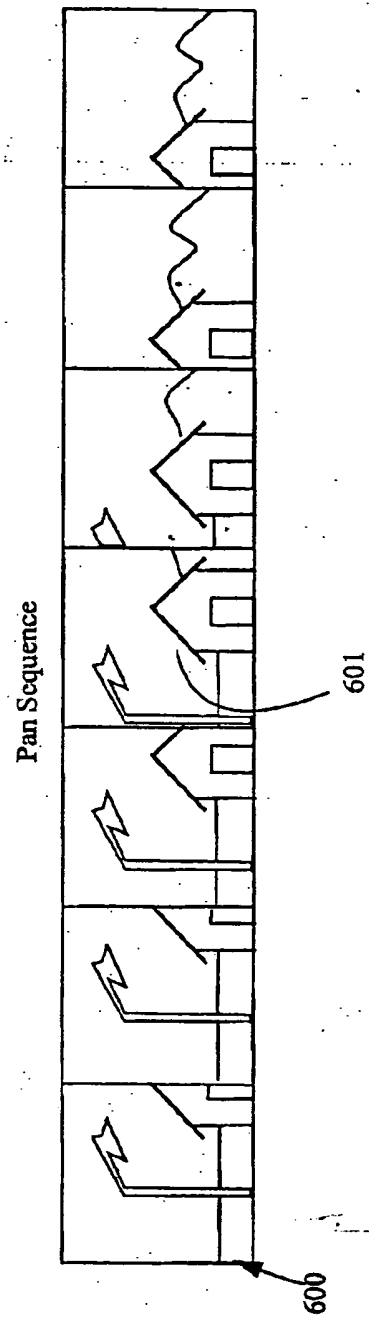


Fig. 10

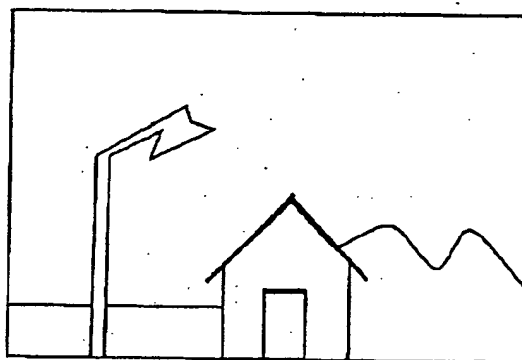


Fig. 11

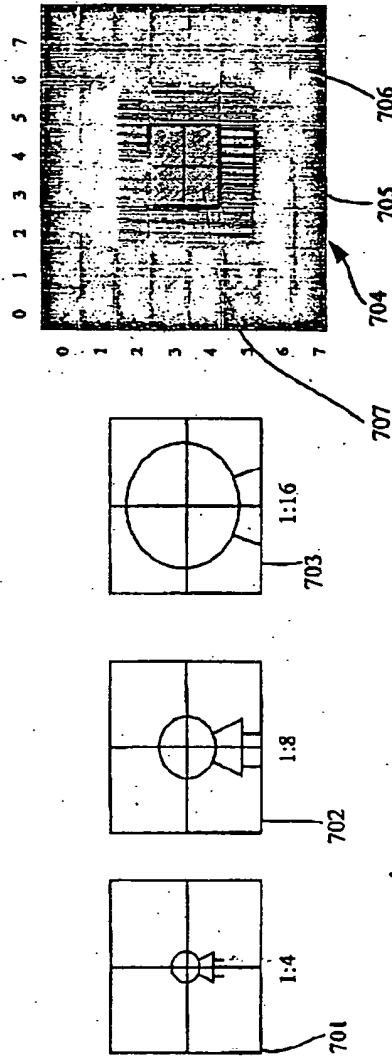


Fig 12

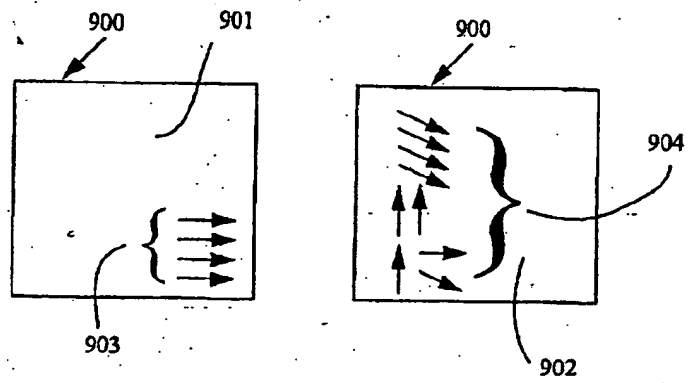


Fig. 13

(100)

特開平11-234560

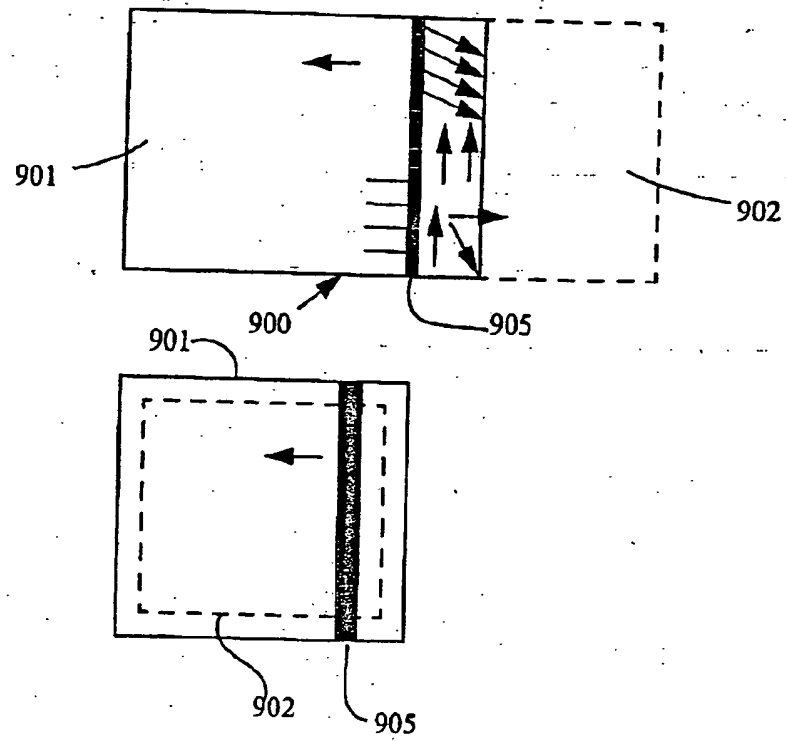


Fig. 14

(101)

特開平11-234560

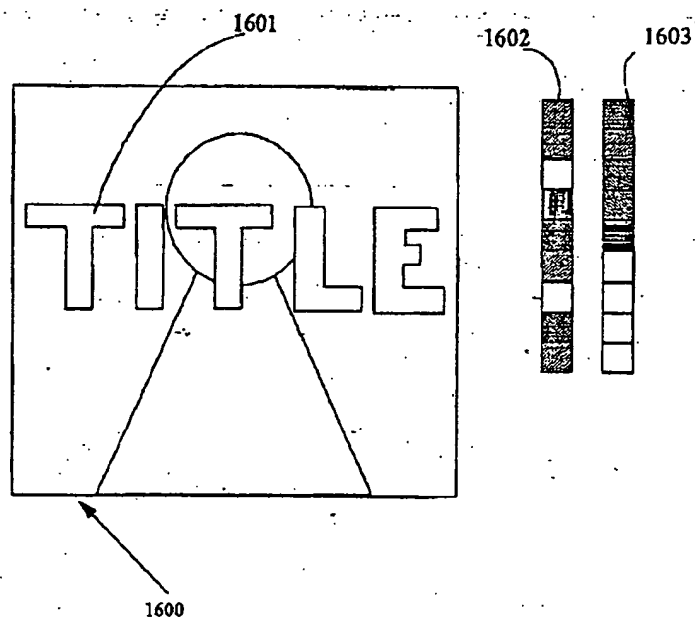


Fig. 15

(102)

特開平11-234560

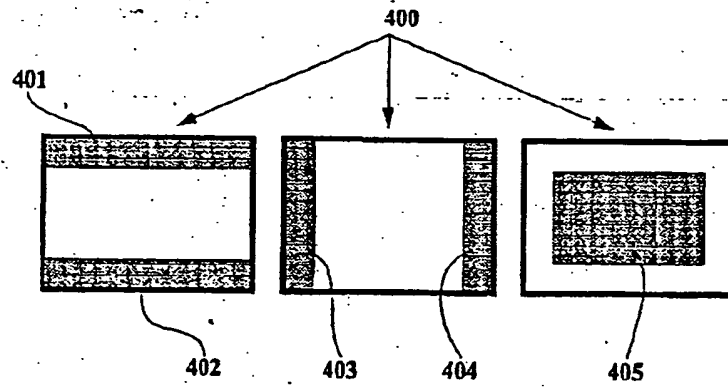


Fig. 16



(103)

特開平11-234560

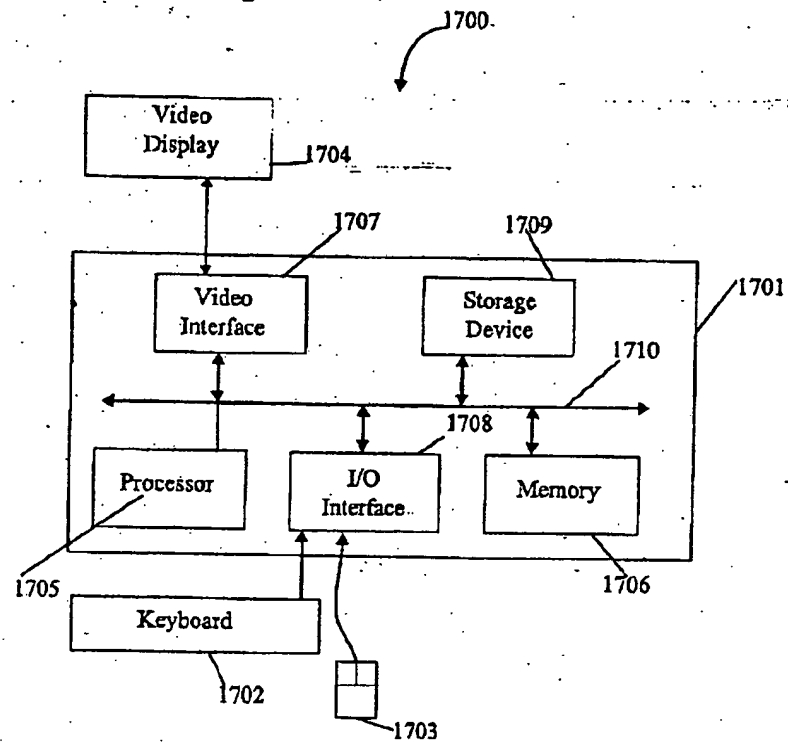


Fig. 17

(104)

特開平11-234560

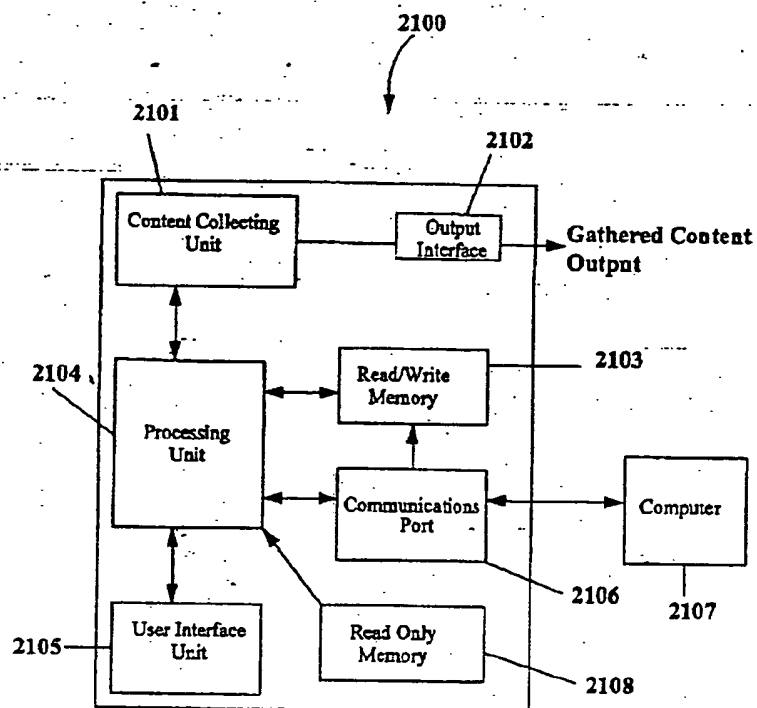


Fig. 18

(105)

特開平11-234560

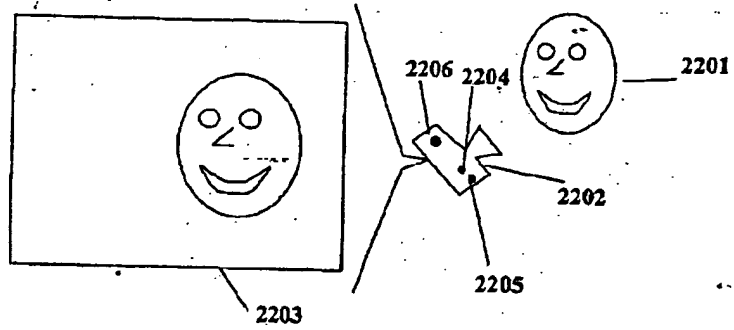


Fig. 19

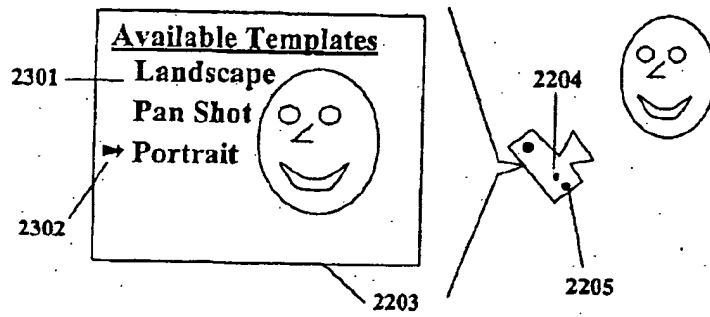


Fig. 20

(106)

特開平11-234560

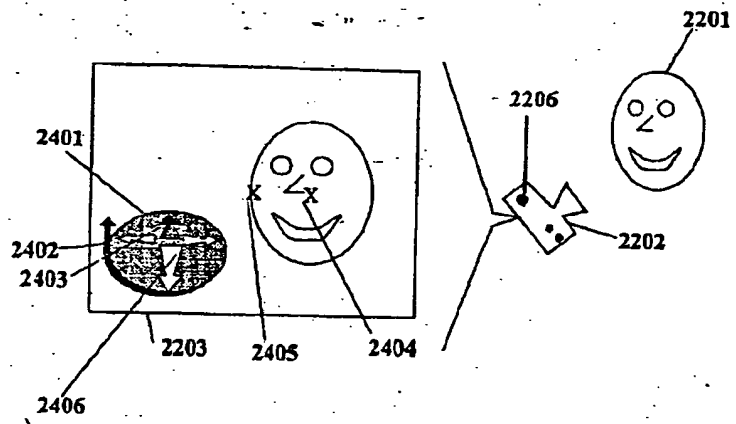


Fig. 21

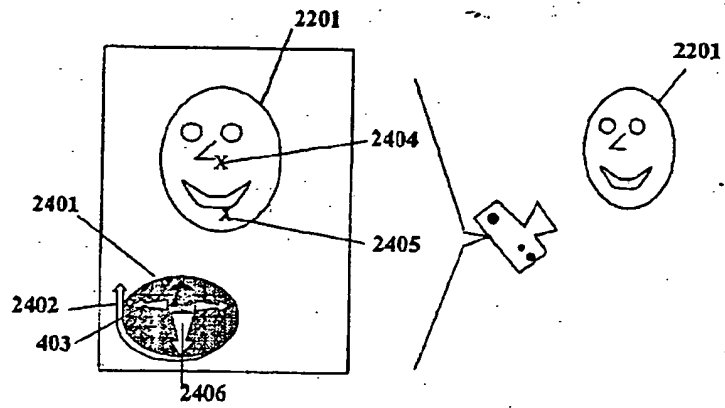


Fig. 22

(107)

特開平11-234560

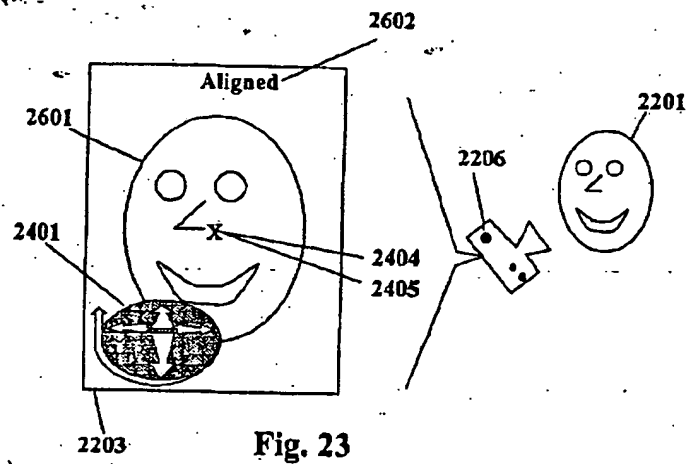


Fig. 23

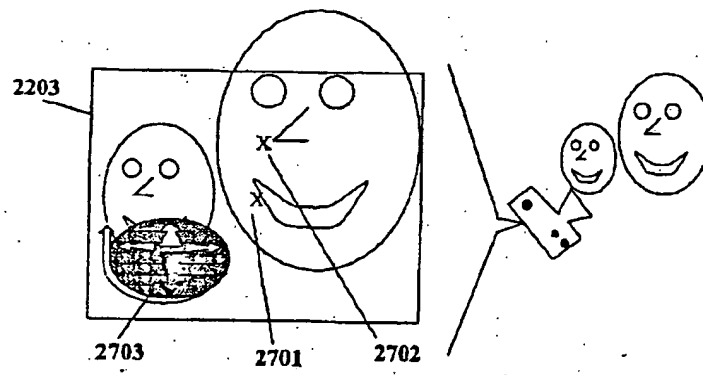


Fig. 24

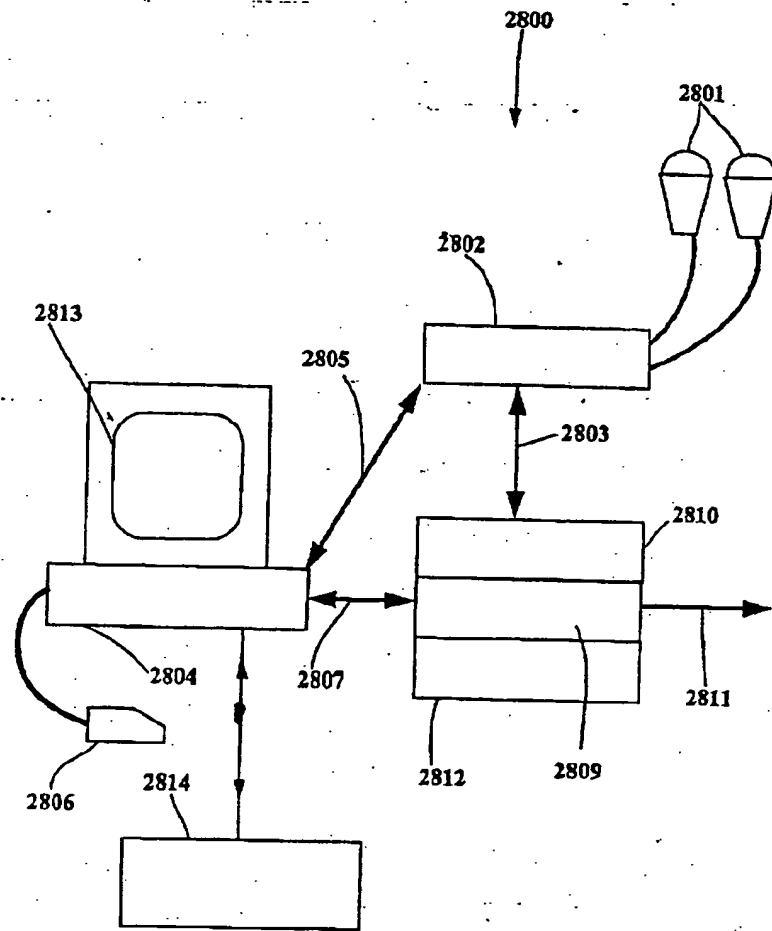


Fig. 25

2903

**REAL ESTATE CHECK LIST**

2900

<input checked="" type="checkbox"/>	Front Garden
<input checked="" type="checkbox"/>	Main Entrance
<input type="checkbox"/>	Kitchen
<input type="checkbox"/>	Bathroom
<input type="checkbox"/>	Laundry
<input type="checkbox"/>	Main Bedroom
<input type="checkbox"/>	Guest Room

2902

2901

Fig. 26

(110)

特開平11-234560

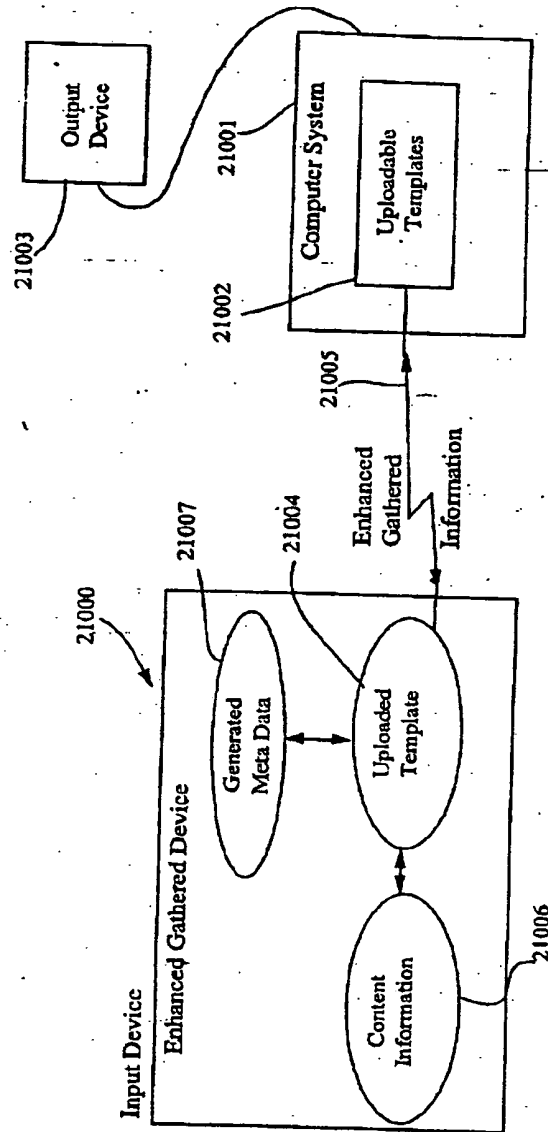


Fig. 27



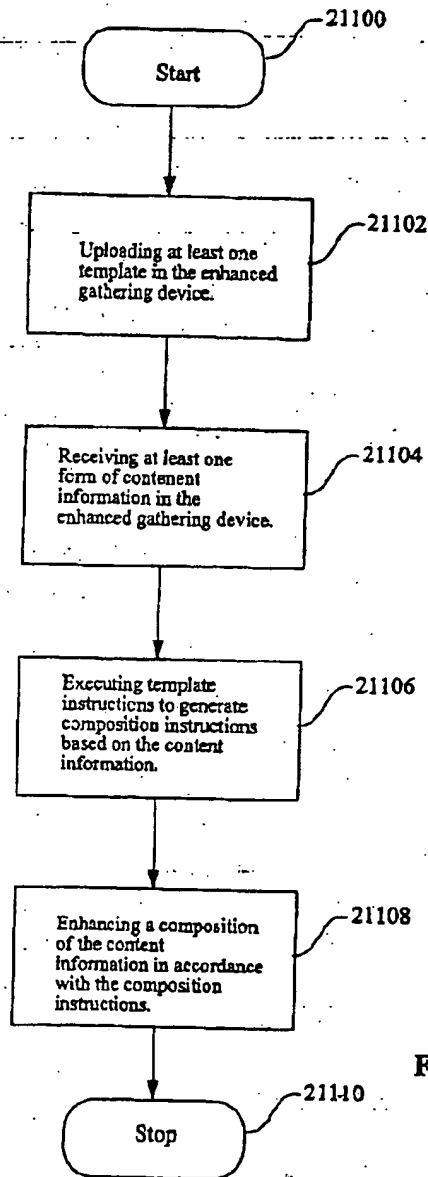


FIG. 28

**1 Abstract**

A method of forming a media production from source media including at least one of audio media and visual media is disclosed. The source media is received and analysed to extract therefrom content information associated with a corresponding element of the source media. The content information is examined for at least one of the elements to determine a limited set of production rules applicable to the corresponding element. The user may selectively apply at least one of the rules to a corresponding one of the elements so as to alter the corresponding element, for example in duration, synchronicity with other elements etc. The altered element is then combined at least with other of the media to form the media production.

**2 Representative Drawing**

Fig. 1

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**This Page Blank (uspto)**